

Selected Problems in Enumerative  
Combinatorics: Permutation Classes, Random  
Walks and Planar Maps.

Andrew Elvey Price  
(ORCID Number: 0000-0003-3240-6390)

Submitted in total fulfilment the degree *Doctor of Philosophy*

September, 2018

School of Mathematics and Statistics, The University of  
Melbourne

## Abstract

In this thesis we consider a number of enumerative combinatorial problems. We solve the problems of enumerating Eulerian orientations by edges and quartic Eulerian orientations counted by vertices. We also find and prove an algebraic relationship between the counting functions for permutations sortable by a double ended queue (deque) and permutations sortable by two stacks in parallel (2sip). In each of these cases, our proof of the result uses an elaborate system of functional equations which is much more complicated than the result itself, leaving the door open for a more direct, combinatorial proof.

We find polynomial time algorithms for generating the counting sequence of deque-sortable permutations and the cogrowth sequence of some groups, including the lamplighter group  $L$  and the Brin-Navas group  $B$ . For permutations sortable by two stacks in series and for the cogrowth sequence of Thompson's group  $F$ , we find exponential time algorithms which are significantly more efficient than the algorithms that previously existed in the literature. In each case an empirical analysis of the produced terms of the sequence leads to a prediction regarding its asymptotic form. In particular, this method leads us to conjecture that the growth rate of deque-sortable permutations is equal to that of 2sip-sortable permutations, a conjecture which we reduce to three conjectures of Albert and Bousquet-Mélou about quarter plane walks. The analysis of the cogrowth sequence of Thompson's group  $F$  leads us to conjecture that  $F$  is not amenable.

We also study the enumeration of 1324-avoiding permutations, a notoriously difficult problem in the field of pattern avoiding permutations. Using a structural decomposition of these permutations, we improve the lower and upper bounds on the growth rate to 10.271 and 13.5 respectively.

Next we investigate the concept of combinatorial Stieltjes moment sequences. We prove that the counting sequence of returns in any undirected locally-finite graph is a Stieltjes moment sequence. As a special case, this implies that any cogrowth sequence is a Stieltjes moment sequence. Based on empirical evidence, we conjecture that the counting sequence for 1324-avoiding permutations is a Stieltjes moment sequence, which would imply an improved lower bound of 10.302 on its growth rate.

We then describe a general class of counting sequences of augmented perfect matchings, which we prove to be Stieltjes moment sequences. In fact, we prove the stronger result that these sequences are Hankel totally-positive

as sequences of polynomials. As a special case, we show that the Ward polynomials are Hankel totally-positive.

In the final chapter we generalise an identity of Duminil-Copin and Smirnov for the  $O(n)$  loop model on the hexagonal lattice to the off-critical case. In the  $n = 0$  case, which corresponds to the enumeration of self-avoiding walks, we use our identity to prove a relationship between the half-plane surface critical exponents  $\gamma_1$  and  $\gamma_{11}$  and the exponent characterising the winding angle distribution of self-avoiding walks in the half-plane.

## **Declaration**

This is to certify that:

- This thesis comprises only my own original work towards the Doctor of Philosophy except where indicated in the Preface.
- Due acknowledgement has been made in the text to all other material used.
- The thesis is fewer than the 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Andrew Elvey Price

## Preface

This thesis draws on a number of the author's papers which were joint work with other researchers. The introduction to power series in Section 1.1 is mostly taken from [49], which is joint work with Mireille Bousquet-Mélou. The original version of this section was written by Bousquet-Mélou, but it has been edited to introduce the language used throughout the thesis.

In Chapter 2, Section 2.3 is mostly taken from [93], and Section 2.4 is mostly taken from [94], which are both joint papers with Tony Guttmann, however these sections are my own work. Section 2.5 is a truncated version of Guttmann's contribution to these two papers.

Chapter 3 is taken from the paper [28], with an extended introduction to introduce pattern avoiding permutations. This paper is joint work with David Bevan, Robert Brignall and Jay Pantone. Sections 3.2, 3.3 and 3.4 were contributions of the other three authors, while the following sections were each about 50% my own contribution. So, overall I contributed 25% of this chapter. The final version of these sections was written by David Bevan.

Chapter 4 is partly taken from [91], which is joint work with Tony Guttmann, and is partly taken from [49], which is joint work with Mireille Bousquet-Mélou. A truncated version of Guttmann's contribution is given in Section 4.3.5. Sections 4.3.1, 4.3.2 and 4.3.3 summarise my own contribution to [91]. All of Section 4.3 is edited to use the same notation as the rest of the chapter. All other sections are joint work with Bousquet-Mélou, and we each contributed 50%.

Chapter 5 is taken from [92], which is joint work with Tony Guttmann. A truncated version of Guttmann's contribution is given in Section 5.5, while all other sections are my own work.

Chapter 6 is my own work. This includes Section 6.1 and 6.2.1, which are largely taken from [92].

Chapter 7 is joint work with Alan Sokal. We each contributed 50%.

Chapter 8 is taken from [90], which is joint work with Jan de Gier, Tony Guttmann and Alexander Lee and was carried out prior to enrolment in the degree. We each contributed 25%.

During my candidature I was financially supported by:

- An Australian government research training program scholarship throughout my candidature.
- An Elizabeth and Vernon Puzey scholarship throughout my candidature.

- The ARC Centre of Excellence for Mathematics and Statistics of Complex Systems (MASCOS) in the form of a top up scholarship between October 2015 and December 2016.
- ACEMS in the form of two travel stipends and a top up scholarship between January 2017 and September 2018.
- The French Agence Nationale de la Recherche, via grant Graal ANR-14-CE25-0014 while in France between December 2017 and May 2018.
- A 2017 Nicolas Baudin travel grant awarded by the French embassy in Australia.

## Acknowledgements

I would like to start by thanking my supervisor Tony Guttmann. Tony introduced me to machine sortable permutations, map enumeration, enumeration of cogrowth sequences and self-avoiding walks, and in each of these areas our collaboration resulted in a paper. We had many interesting discussions on all of the topics discussed in this thesis. I am grateful to him for introducing me to Mireille Bousquet-Mélou, Robert Brignall and Alan Sokal by arranging for me to visit each of them in Europe. I also thank him for helping me to attend a variety of national and international conferences. Tony taught me a lot about mathematics, especially numerical analysis. He also taught me a great deal about writing and dealing with administration.

I am indebted to Mireille Bousquet-Mélou, who hosted me in Bordeaux for two weeks in 2017 then another 6 months in 2017-2018. During this time she taught me a lot about map enumeration, functional equations and how to write a good paper. Some of our discussions were the most fruitful of my studies, resulting in a complete solution for the enumeration of Eulerian orientations. I am also grateful to Mireille for encouraging me to attend a number of conferences and present my research while I was in Europe.

I would like to thank Robert Brignall, Alan Sokal and Kilian Raschel, all of whom hosted me in Europe during my PhD studies. During the week I spent with Robert Brignall we had many fruitful discussions about his work in progress with David Bevan and Jay Pantone, resulting in a paper between all four of us. While I was hosted by Alan Sokal, he taught me a lot about continued fractions, Stieltjes moment sequences and Hankel total positivity. This has resulted in a number of works in progress between us, one of which is contained in this thesis. I thank Kilian Raschel for teaching me about quarter plane walks and the analytic methods that he has used to solve functional equations describing their generating functions.

I am grateful to my other coauthors David Bevan, Jay Pantone, Jan de Gier and Alex Lee who were all very careful to write our papers in the clearest possible way.

I would also like to thank a number of other researchers with whom I had interesting discussions about topics mentioned in this thesis: Matthew Brin, Nicolas Bonichon, Jeremy Bouttier, Nathan Clisby, Andrew Conway, Murray Elder, Jean-Marie Maillard, Maria Ramirez-Solano, Andrew Rechnitzer, Michael Wallner and Paul Zinn-Justin.

Finally I want to thank all of my friends and family for their encouragement and moral support during the last three years.

# Contents

<b>1</b>	<b>Introduction</b>	<b>11</b>
1.1	Formal power series . . . . .	18
<b>2</b>	<b>Machine sortable permutations</b>	<b>20</b>
2.1	Introduction . . . . .	20
2.2	Permutations sortable by $m$ stacks in parallel . . . . .	22
2.2.1	Notation, operation sequences and perfect matchings . . . . .	22
2.3	Permutations sortable by a double ended queue . . . . .	27
2.3.1	Notation and canonical operation sequences . . . . .	27
2.3.2	Enumeration . . . . .	33
2.3.3	Analysis . . . . .	43
2.3.4	Algorithms for 2sip and deque-sortable permutations . . . . .	48
2.4	Permutations sortable by two stacks in series . . . . .	50
2.4.1	Basic algorithm . . . . .	50
2.4.2	Forbidden words and regular languages . . . . .	51
2.4.3	Increment-avoiding permutations . . . . .	54
2.4.4	Memory consumption and parallelisation . . . . .	55
2.4.5	Results . . . . .	55
2.5	Empirical analysis using differential approximants . . . . .	56
2.5.1	Ratio method. . . . .	56
2.5.2	Method of differential approximants . . . . .	57
2.5.3	Analysis of series for deque and 2sip . . . . .	58
2.5.4	Series extension for 2sis. . . . .	63
2.5.5	Analysis of extended series for 2sis . . . . .	66
2.6	Improved lower bounds for permutations sortable by 2sip, 2sis and a deque . . . . .	70
<b>3</b>	<b>Bounds on the growth rate of 1324-avoiding permutations</b>	<b>72</b>
3.1	Background on pattern avoiding permutations . . . . .	72
3.2	The growth rate of 1324-avoiding permutations . . . . .	74
3.3	Staircase structure . . . . .	75



3.4	1324-avoiding dominoes . . . . .	79
3.4.1	Arch systems . . . . .	80
3.4.2	Arch configurations . . . . .	82
3.4.3	The enumeration of dominoes . . . . .	85
3.4.4	Balanced dominoes . . . . .	87
3.5	An upper bound . . . . .	87
3.6	An initial lower bound . . . . .	89
3.7	Domino substructure . . . . .	92
3.7.1	Leaves . . . . .	92
3.7.2	Empty strips . . . . .	96
3.7.3	Dominoes with many leaves and many empty strips . . . . .	98
3.8	A better lower bound . . . . .	99
3.8.1	Horizontally interleaved connecting cells . . . . .	101
3.8.2	Refining the staircase . . . . .	105
3.8.3	Enumerating the refined staircase . . . . .	109
3.8.4	Improving the lower bound further . . . . .	110
<b>4</b>	<b>Enumerating planar Eulerian orientations</b>	<b>112</b>
4.1	Introduction . . . . .	112
4.2	Definitions . . . . .	116
4.2.1	Planar maps . . . . .	116
4.2.2	Orientations . . . . .	117
4.2.3	Eulerian orientations, the 6-vertex model and fully packed loops . . . . .	119
4.3	Numerical work . . . . .	122
4.3.1	Contraction Operation . . . . .	122
4.3.2	Functional equations for quartic Eulerian orientations . . . . .	124
4.3.3	Functional equations for general Eulerian orientations . . . . .	127
4.3.4	Algorithmic computation of generating functions . . . . .	131
4.3.5	Empirical analysis of generating functions . . . . .	131
4.4	Conjecturing exact forms of $Q(t)$ and $G(t)$ . . . . .	137
4.5	Simplified functional equations for quartic Eulerian orientations	139
4.6	Solution for quartic Eulerian orientations . . . . .	150
4.7	A bijection . . . . .	155
4.7.1	From labelled maps to mobile-maps . . . . .	155
4.7.2	Specialization to labelled quadrangulations . . . . .	158
4.7.3	Specialization to canonically labelled maps . . . . .	160
4.8	Functional equations for general Eulerian orientations . . . . .	161
4.9	Solution for general Eulerian orientations . . . . .	167
4.10	Nature of the series and asymptotics . . . . .	170
4.10.1	Nature of the series . . . . .	170

4.10.2	Asymptotics . . . . .	172
4.10.3	Accuracy of the empirical analysis . . . . .	174
4.11	Final comments and perspectives . . . . .	174
4.11.1	Bijections . . . . .	174
4.11.2	Interpolating between quartic Eulerian orientations and general Eulerian orientations . . . . .	178
<b>5</b>	<b>Enumerating cogrowth numbers in Thompson's group <math>F</math> and related groups</b>	<b>180</b>
5.1	Introduction . . . . .	180
5.2	Preliminaries . . . . .	182
5.3	Series generation . . . . .	186
5.3.1	Wreath Products $G \wr \mathbb{Z}$ . . . . .	186
5.3.2	The Brin-Navas group $B$ . . . . .	190
5.3.3	A General Algorithm . . . . .	193
5.3.4	Thompson's group $F$ . . . . .	199
5.4	Possible cogrowth of the groups $F$ and $B$ . . . . .	200
5.5	Empirical analysis using differential approximants . . . . .	203
5.5.1	Analysis of the Brin-Navas group $B$ . . . . .	203
5.5.2	Analysis of Thompson's group $F$ . . . . .	206
5.6	Conclusion . . . . .	209
<b>6</b>	<b>Combinatorial Stieltjes moment sequences</b>	<b>212</b>
6.1	Introduction to Stieltjes moment sequences . . . . .	212
6.1.1	Bounding the growth rate of Stieltjes moment sequences	216
6.2	Methods for proving that combinatorial sequences are Stieltjes moment sequences . . . . .	217
6.2.1	Loops in graphs and cogrowth sequences . . . . .	217
6.2.2	Exactly solved sequences . . . . .	219
6.2.3	Continued fraction approach . . . . .	222
6.2.4	Permutations avoiding increasing patterns . . . . .	223
6.3	Empirical Stieltjes moment sequences . . . . .	224
6.3.1	Empirical analysis of known Stieltjes moment sequences	224
6.3.2	Pattern avoiding permutations . . . . .	227
6.3.3	Eulerian orientations . . . . .	229
6.3.4	Machine sortable permutations . . . . .	230
6.4	Hankel total positivity . . . . .	232

<b>7</b>	<b>Phylogenetic trees, augmented perfect matchings and a Thron-type continued fraction (T-fraction) for the Ward polynomials</b>	<b>233</b>
7.1	Introduction . . . . .	233
7.2	Statement of results . . . . .	245
7.2.1	S-fractions for perfect matchings . . . . .	245
7.2.2	T-fractions for super-augmented perfect matchings . . . . .	249
7.3	Preliminaries . . . . .	254
7.3.1	Combinatorial interpretation of continued fractions . . . . .	254
7.3.2	Labelled Schröder paths . . . . .	256
7.4	Proof of Theorem 7.3 . . . . .	257
7.5	Bijection between phylogenetic trees and augmented perfect matchings . . . . .	261
7.6	Recurrence for polynomials defined by the general linear T-fraction . . . . .	263
<b>8</b>	<b>Off-critical parafermions and the winding angle distribution of the <math>O(n)</math> model</b>	<b>266</b>
8.1	Introduction . . . . .	266
8.2	Off-critical identity for the honeycomb $O(n)$ model . . . . .	267
8.2.1	Off-critical deformation . . . . .	270
8.3	Winding angle . . . . .	273
8.3.1	Susceptibilities and critical exponents . . . . .	274
8.3.2	Asymptotic winding angle distribution . . . . .	275
8.3.3	Winding angle in a wedge . . . . .	276
8.3.4	Exponent inequalities . . . . .	277
8.4	Conjectures . . . . .	278
8.4.1	Winding angle distribution from conformal field theory . . . . .	278
8.4.2	Wedge exponents . . . . .	279
8.5	Conclusion . . . . .	281

# List of Figures

2.1	The input and output operations on a stack. . . . .	20
2.2	Example of a 3-coloured perfect matching. . . . .	24
2.3	Example of a 3-colourable, reduced perfect matching. . . . .	26
2.4	The input and output operations on a deque. . . . .	28
2.5	Example of a procedure on a deque producing the permutation 4123. . . . .	28
2.6	Two stacks in series . . . . .	50
2.7	The infinite state automaton $\Gamma$ . . . . .	53
2.8	Exponent estimates $\theta_n$ plotted against $1/n$ . . . . .	59
2.9	Exponent estimates $g_n$ plotted against $1/n$ for deques. . . . .	60
2.10	Extrapolated exponent estimates $h_n$ plotted against $1/n$ for deques. . . . .	60
2.11	Exponent estimates $g_n$ plotted against $1/n$ for two stacks in parallel. . . . .	62
2.12	Extrapolated exponent estimates $g_n$ plotted against $1/n$ for two stacks in parallel. . . . .	62
2.13	Plot of ratios against $1/n$ for 2sis. . . . .	68
2.14	Plot of intercepts of successive ratios against $1/n^2$ for 2sis. . .	69
2.15	Plot of estimators of exponent $g$ against $1/n^2$ for 2sis for dif- ferent estimates of $\mu$ . . . . .	70
3.1	The descending $(\text{Av}(213), \text{Av}(132))$ staircase containing $\text{Av}(1324)$ . .	76
3.2	The greedy gridding of a 1324-avoider in the staircase. . . . .	77
3.3	The greedy gridding of a 1324-avoider of length 1000. . . . .	78
3.4	Four distinct small dominoes. . . . .	79
3.5	A 213-avoider and a 132-avoider with their arch systems. . . .	81
3.6	Mapping an arch system to a 213-avoider. . . . .	82
3.7	The arch configuration for a domino. . . . .	82
3.8	The six ways of decomposing a non-empty arch prefix in $\mathcal{A}$ . . .	84
3.9	Mapping a greedy-gridded 1324-avoider to a binary word and a domino. . . . .	88

3.10	The decomposition of the staircase into dominoes and connecting cells. . . . .	89
3.11	Interleaving the skew indecomposable components in a connecting cell with the points in the two adjacent domino cells. .	90
3.12	Strips in a domino cell. . . . .	96
3.13	The decomposition of the staircase into dominoes and connecting cells. . . . .	100
3.14	Interleaving the points in a connecting cell with those in two domino cells. . . . .	101
3.15	The scheme used to calculate the improved lower bound on 1324-avoiders. . . . .	105
4.1	Example of a quartic planar Eulerian orientation. . . . .	113
4.2	The planar Eulerian orientations with at most two edges. . . .	114
4.3	A rooted planar map and its dual map. . . . .	117
4.4	A labelled map. . . . .	118
4.5	An Eulerian orientation and the corresponding dual labelled map. . . . .	119
4.6	The two types of vertices in a quartic Eulerian orientation. . .	120
4.7	Transformation from vertices in a quartic Eulerian orientation to pairs of vertices in certain cubic Eulerian partial orientations.	122
4.8	Example of a minus subpatch and the corresponding contraction operation. . . . .	123
4.9	Example of a patch and D-patch. . . . .	125
4.10	Ratio plot of coefficients of $G(t)$ . . . . .	133
4.11	Ratio plot of coefficients of $Q(t)$ . . . . .	133
4.12	Plot of linear intercepts of ratios of $G(t)$ vs. $1/n \log^2 n$ . . . .	133
4.13	Plot of linear intercepts of ratios of $Q(t)$ vs. $1/n \log^2 n$ . . . .	133
4.14	Plot of linear intercepts of ratios of $Q(t)$ vs. $1/n \log^2 n$ , using an extra 1000 ratios. . . . .	134
4.15	Plot of linear intercepts of ratios of $Q(t)$ vs. $1/n \log^2 n$ , using ratios 700 to 1100. . . . .	134
4.16	Plot of linear intercepts of ratios of $G(t)$ vs. $1/n \log^2 n$ , using an extra 1000 ratios. . . . .	135
4.17	Plot of linear intercepts of ratios of $G(t)$ vs. $1/n \log^2 n$ , using ratios 700 to 1100. . . . .	135
4.18	Plot of exponent $\alpha$ estimates from $G(t)$ vs. $1/\log n$ , using an extra 1000 ratios. . . . .	136
4.19	Plot of exponent $\alpha$ estimates from $Q(t)$ vs. $1/\log n$ , using an extra 1000 ratios. . . . .	136
4.20	Plot of exponent $\alpha$ estimates from eqn (4.4). . . . .	137

4.21	Plot of exponent $\beta - 1$ estimates from eqn (4.4).	137
4.22	Plot of exponent $\beta$ estimates from eqn (4.5).	138
4.23	C-patch and D-patch examples.	141
4.24	The transformation of $C$ into $P$ used to prove Lemma 4.17.	142
4.25	A sequence of C-patches and the corresponding B-patch.	142
4.26	The transformation from a B-patch to a D-patch.	143
4.27	The four different types of patches which contribute to the generating function $[y^1]\mathbf{C}(t, x, y)$ .	143
4.28	Example of minus-subpatch and the corresponding contraction operation.	145
4.29	Reverse of the contraction operation.	147
4.30	The three different types of D-patches.	147
4.31	The transformation of a labelled map $L$ into a D-patch $D'$ .	148
4.32	A mobile-map with 2 faces.	156
4.33	Examples of the operation $\Phi$ .	156
4.34	A labelled map $L$ and the associated mobile-map $M = \Phi(L)$ .	157
4.35	Transformations from a labelled map to a colourful labelled quadrangulation.	160
4.36	Transformation from a quartic Eulerian orientation to an Eulerian <i>partial</i> orientation.	161
4.37	The map $\Phi$ applied to a canonically labelled bipartite map.	162
4.38	A colourful C-patch $C$ with outer degree 2 and the corresponding patch $P$ .	164
4.39	Transformation from a C-patch $C$ to a shifted patch $S$ and D-patch $D$ .	165
4.40	An example of the transformation from a colourful C-patch $C$ to a colourful labelled quadrangulation $C'$ .	168
4.41	Small quartic graph and the corresponding numbers of quartic planar maps and spanning forests.	176
4.42	The trees with 2 and 3 white leaves counted by $t - R(t)$ .	176
4.43	The trees with at most 4 leaves involved in the expansion of $t - R(t)$ .	177
5.1	The first 128 modified ratios of the Brin-Navas group $B$ vs. $1/n$ .	204
5.2	The first 128 modified ratios for the Brin-Navas group $B$ vs. $1/\log n$ .	205
5.3	The first 718 modified ratios for the Brin-Navas group $B$ vs. $1/\log n$ .	205
5.4	Estimates of $\sigma - 2$ from 128 terms of the Brin-Navas group $B$ .	206
5.5	Estimates of $\sigma - 2$ from 256 terms of the Brin-Navas group $B$ .	206

5.6	Modified ratios vs. $1/n$ for Thompson's group $F$ .	207
5.7	Modified ratios vs. $n^{-1/5}$ for Thompson's group $F$ .	207
5.8	Estimators of $\sigma - 2$ for Thompson's group $F$ vs. $1/n$ .	208
5.9	Estimators of $-\delta$ for Thompson's group $F$ vs. $1/n$ .	208
5.10	The first 186 modified ratios for Thompson's group $F$ plotted against $1/\sqrt{n}$ .	208
5.11	The first 186 modified ratios for Thompson's group $F$ plotted against $\sqrt{\log n/n}$ .	208
5.12	Quotient of $F$ and $L$ ratios using 200 terms.	210
5.13	Quotient of $F$ and $\mathbb{Z} \wr \mathbb{Z}$ ratios using 200 terms.	210
5.14	Quotient of $F$ and $(\mathbb{Z} \wr \mathbb{Z}) \wr \mathbb{Z}$ ratios using 200 terms.	210
6.1	Plot of $\alpha_n$ vs. $1/n$ for the sequence $\text{Av}_n(1342)$ , using $n \in [10, 50]$ .	225
6.2	Plot of $\alpha_n$ vs. $1/n$ for the sequence $\text{Av}_n(1234)$ , using $n \in [10, 50]$ .	225
6.3	Plot of $\alpha_n$ vs. $1/n$ for the cogrowth sequence of Heisenberg's group, using $n \in [5, 50]$ .	226
6.4	Plot of $\alpha_n$ vs. $\log(n)/n$ for the cogrowth sequence of $\mathbb{Z} \wr \mathbb{Z}$ , using $n \in [5, 50]$ .	226
6.5	Plot of $\alpha_n$ vs. $\log(n)^{2/3}n^{-2/3}$ for the cogrowth sequence of $\mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z}$ , using $n \in [5, 50]$ .	226
6.6	Plot of $\alpha_n$ vs. $1/n$ for the cogrowth sequence of the Brin-Navas group $B$ , using $n \in [5, 50]$ .	227
6.7	Plot of $\alpha_n$ vs. $1/n$ for the cogrowth sequence of Thompson's group $F$ , using $n \in [5, 31]$ .	227
6.8	Plot of $\alpha_n$ vs. $n^{-2/3}$ for the sequence $\text{Av}_n(1324)$ , using $n \in [10, 50]$ .	228
6.9	Plot of $\alpha_n$ vs. $1/n^2$ for the sequence $q_n$ enumerating quartic Eulerian orientations, using $n \in [20, 50]$ .	229
6.10	Plot of $\alpha_n$ vs. $1/n$ for the sequence $g_n$ enumerating Eulerian orientations, using $n \in [10, 50]$ .	229
6.11	Plot of estimates of $\mu$ vs. $1/n^2$ for the sequence $\mathbf{g}$ , using $n \in [50, 100]$ .	230
6.12	Plot of $\alpha_n$ vs. $1/n$ for the sequence $p_n$ enumerating 2sip-sortable permutations, using $n \in [10, 50]$ .	231
6.13	Plot of $\alpha_n$ vs. $1/n$ for the sequence $d_n$ enumerating deque-sortable permutations, using $n \in [10, 50]$ .	231
6.14	Plot of $\alpha_n$ vs. $1/n$ for the sequence $s_n$ enumerating 2sis-sortable permutations, using $n \in [5, 19]$ .	231
7.1	Examples of wiggly and dashed lines in an augmented perfect matching.	237

7.2	Example of a super-augmented perfect matching and the corresponding labelled Schröder path. . . . .	259
7.3	Example of the bijection between super-augmented perfect matchings and phylogenetic trees. . . . .	262
7.4	The five cases in the proof of Proposition 7.9. . . . .	265
8.1	A configuration $\gamma$ on a finite domain. . . . .	268
8.2	The two ways of grouping the configurations which end at mid-edges $p, q, r$ adjacent to vertex $v$ . . . . .	269
8.3	The three possible ways for a walk to enter a given vertex via each of the three mid-edges, $p, q$ and $r$ . . . . .	272
8.4	A walk terminating at the mid-edge $z$ . . . . .	273
8.5	Finite patch $S_{3,1}$ of the hexagonal lattice. . . . .	274



# List of Tables

1.1	A table summarising the current state of some of the enumerative problems considered in this thesis. . . . .	17
1.2	A table summarising the estimated asymptotic form of some of the sequences considered in this thesis. . . . .	18
2.1	Predicted coefficients $p_{20}$ to $p_{38}$ for two stacks in parallel with estimated and exact error. . . . .	64
2.2	Predicted coefficients $s_{20}$ to $s_{38}$ for two stacks in series with estimated error. . . . .	65
2.3	Predicted ratios $r_n = s_n/s_{n-1}$ and their standard deviation for two stacks in series. . . . .	67
3.1	A chronology of lower and upper bounds for $\text{gr}(\text{Av}(1324))$ . . .	74
5.1	The first 50 coefficients of the cogrowth series for the Brin-Navas group. . . . .	194
5.2	Terms in the cogrowth sequence of Thompson's group $F$ . . . .	201

# Chapter 1

## Introduction

Enumerative combinatorics is concerned with questions of the form “how many objects of size  $n$  are there”. Any such problem defines a counting sequence  $\mathbf{a} = a_0, a_1, a_2, \dots$  which one can study. An ideal result would be to find an exact closed form solution for the terms in this sequence, however this is not always possible. Instead we might find a closed form for the ordinary generating function  $A(t) = \sum_{n=0}^{\infty} a_n t^n$  or exponential generating function  $\bar{A}(t) = \sum_{n=0}^{\infty} \frac{a_n}{n!} t^n$  of such a sequence, or perhaps a way to characterise the generating function as the solution of a differential equation or a system of functional equations. Any of these solutions will generally result in a polynomial time algorithm for computing the terms of the sequence. Sometimes even that is beyond our reach, and we instead search for faster exponential time algorithms for computing the terms of the sequence.

Apart from exact enumeration, we are also interested in the asymptotic behaviour of such a counting sequence  $\mathbf{a}$ , and particularly in the growth rate

$$\mu = \lim_{n \rightarrow \infty} \sqrt[n]{a_n},$$

when it exists. When  $\mu$  exists, one often has the form

$$a_n \sim c \cdot \mu^n n^g,$$

for some constants  $c$  and  $g$ . We also study some sequences with more complicated asymptotic forms, for example including logarithmic terms  $\log(n)^k$  as in Chapter 4 or stretched exponential terms  $\rho^{n^\kappa}$  as in chapters 3 and 5. Understanding the asymptotics in this way allows us to approximate terms  $a_n$  which are well beyond our ability to calculate exactly. Of course, when we have an exact, closed form solution, it generally allows us to deduce the asymptotic behaviour of the sequence. This is also often possible when we have a differential equation describing the associated generating function,

by applying methods from [106]. In other cases we use two approaches - on one hand we can seek to rigorously prove bounds on the terms  $a_i$  and hence the growth rate  $\mu$ , by finding subclasses and superclasses of our combinatorial class which we can enumerate exactly, as we do in Chapter 3 for 1324-avoiding permutations. On the other hand we can empirically analyse the terms of the sequence that we have calculated exactly to estimate the asymptotic behaviour. We present Guttmann's empirical analysis of some sequences in Sections 2.5, 4.3.5 and 5.5. In each case, Guttmann's method of differential approximants [126, 125, 123, 124] is used.

The first problems that we consider in this thesis are those of enumerating permutations that are sortable using certain classical data structures from computer science. In particular, we consider the problems of sorting by two stacks in parallel (2sip), a double ended queue (deque) and two stacks in series (2sis). These three problems were posed by Knuth in his seminal book "The Art of Computer Science" [141], following his solution to the problem of sorting by a single stack. In 2010 all three problems were studied by Albert, Atkinson and Linton [2] who proved upper and lower bounds on their respective exponential growth rates. In 2015, Albert and Bousquet-Mélou were able to characterise the generating function  $P(t)$  of 2sip-sortable permutations using a system of functional equations, thereby yielding a polynomial time algorithm for computing the coefficients of  $P(t)$ . We adapt their method to the case of enumerating deque-sortable permutations, and, in doing so, we derive the following algebraic relationship between the generating function  $D(t)$  of deque-sortable permutations and the generating function  $P(t)$  of 2sip-sortable permutations:

$$2D(t) = 2 + t + 2Pt - 2Pt^2 - t\sqrt{1 - 4P + 4P^2 - 8P^2t + 4P^2t^2 - 4Pt}.$$

Using this equation, we derive that, subject to three conjectures of Albert and Bousquet-Mélou, the generating functions  $P(t)$  and  $D(t)$  have the same radius of convergence. We then use the equations to produce over 1000 coefficients of each of the generating functions  $P(t)$  and  $D(t)$ . Guttmann's analysis of these coefficients, which we summarise in Section 2.5, strongly supports the conjecture that  $P(t)$  and  $D(t)$  have the same radius of convergence, which is  $1/8.281402207$ . We then move on to 2sis-sortable permutations, for which we develop an algorithm enabling us to compute the first 20 terms of the counting sequence. Guttmann's analysis of these terms, which we summarise in Section 2.5, suggests that the growth rate is around 12.4.

Chapter 3 is devoted to classical pattern avoiding permutations, and in particular, joint work with Bevan, Brignall and Pantone in which we improve the rigorous bounds on the growth rate of 1324-avoiding permutations. A

permutation  $\pi$  is said to contain a permutation  $p = a_1 a_2 \dots a_k$  if there is some subsequence  $b_1, \dots, b_k$  of  $\pi$  which has the same relative order as  $p$ . That is,  $b_i < b_j$  if and only if  $a_i < a_j$ . In this context, the permutation  $p$  is generally called a pattern. The permutation  $\pi$  is said to *avoid* the pattern  $p$  if it does not contain  $p$ . For example, the permutation 2736514 contains the pattern 4231, since the numbers 7351 have the same relative order as 4231. On the other hand, 2736514 avoids the pattern 1324. The set of permutations  $\pi$  which avoid a pattern  $p$  is denoted  $\text{Av}(p)$ , and the set of those of length  $n$  is denoted  $\text{Av}_n(p)$ . More generally, given some set  $S$  of permutation patterns, the set of permutations which avoid all patterns in  $S$  is denoted  $\text{Av}(S)$ . For each machine  $M$  considered in Chapter 2 (deque, 2sip, 2sis), the set  $P_M$  of  $M$ -sortable permutations is equal to  $\text{Av}(S)$  for some set  $S$  of permutation patterns, however  $S$  must be infinite [171, 160]. In other words,  $P_M$  is a permutation class with an infinite basis.

Bounds on the growth rate  $\text{gr}(\text{Av}(1324))$  of 1324-avoiding permutations have been the focus of a number of papers over the last decade and a half [35, 36, 4, 65, 39, 40, 27]. Our improved bounds rely on a structural characterisation of 1324-avoiders as a staircase of 132-avoiders and 213-avoiders. Using this characterisation, one can immediately deduce a lower bound of 9 and an upper bound of 16 on the growth rate  $\text{gr}(\text{Av}(1324))$ . To improve these bounds we first enumerate a restricted class of these staircases, which we call 1324-avoiding dominoes. Using our enumeration of 1324-avoiding dominoes, our upper bound of 13.5 on  $\text{gr}(\text{Av}(1324))$  follows fairly easily, as does a lower bound of 10.125. Our improved lower bound of 10.271 results from a more detailed analysis of the structure of 1324-avoiding dominoes. We observe that 1324-avoiding dominoes are equinumerous with West 2-stack sortable permutations [213], rooted non-separable planar maps [57] and fighting fish [77, 78, 101], but so far we have no bijective proof of this fact.

In Chapter 4 we study planar Eulerian orientations. A planar map is a connected planar graph embedded in the sphere, taken up to orientation preserving homomorphism. A planar Eulerian orientation is a rooted planar map with oriented edges with the property that the in-degree of each edge is equal to its out-degree. The problem of enumerating planar Eulerian orientations (by edges) was posed in 2016 by Bonichon, Bousquet-Mélou, Dorbec and Pennarun [42], who found a sequence of sub-classes and super-classes of Eulerian orientations, with which they proved that the associated growth rate  $\mu$  exists and lies in the interval  $(11.56, 13.005)$ . We have found the following exact solution to this enumeration problem: Let  $R(t)$  be the unique

power series with constant term 0 satisfying

$$t = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n}^2 R^{n+1}.$$

Then the generating function of planar Eulerian orientations, counted by edges is

$$G(t) = \frac{1}{4t^2}(t - 2t^2 - R(t)).$$

Our solution implies that the growth rate  $\mu$  is exactly  $4\pi$ , which does, of course, lie in the interval  $(11.56, 13.005)$ . Bonichon et al also posed the problem of enumerating *quartic* planar Eulerian orientations, in which each vertex is restricted to having degree exactly 4. In physics, this problem is known as the ice model on a random lattice [143, 214]. Our solution to this problem takes a similar form: Let  $R(t)$  be the unique power series with constant term 0 satisfying

$$t = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n} \binom{3n}{n} R^{n+1}.$$

Then the generating function of quartic planar Eulerian orientations, counted by vertices is

$$Q(t) = \frac{1}{3t^2}(t - 3t^2 - R(t)).$$

It follows from these solutions that  $G(t)$  and  $Q(t)$  are both differentially algebraic series which each satisfy a differential equation of order 2 whose coefficients are polynomials in  $t$ .

In both solutions, the first step is a bijection between planar Eulerian orientations and certain labelled maps. In the general case our solution utilises a bijection of Ambjørn and Budd [7] between labelled maps and certain labelled quadrangulations. For each of these problems we find a structural decomposition of the objects which allows us to characterise the generating function using a system of functional equations. We solve each system of equations using a guess and check approach.

Before finding a complete solution to the two problems of enumerating Eulerian orientations, we did a numerical study of both problems in joint work with Guttmann. In that work we found more complicated systems of functional equations to characterise  $G(t)$  and  $Q(t)$ . We were not able to solve these exactly, but they did allow us to compute around 100 coefficients of each generating function. Guttmann's analysis of these sequences, which we present in Section 4.3.5, led him to conjecture the exact values  $4\pi$  and  $4\sqrt{3}\pi$

of the two growth rates. This led us to conjecture the exact form of each series, which led to the complete solution. This numerical work is described in Section 4.3.

In Chapter 5 we move on to an enumeration problem motivated by geometric group theory, namely the enumeration of cogrowth sequences of groups. The  $n$ th cogrowth number  $a_n$  of a group  $G$  with Cayley graph  $\Gamma$  is defined to be the number of walks of length  $2n$  in  $\Gamma$  which start and end at some fixed vertex  $v_0$ . This is of particular interest in geometric group theory because the value of the growth rate of the cogrowth sequence determines whether the group is “amenable” [120, 67]. We develop polynomial time algorithms for a number of groups which are known to be amenable. These form rare examples of combinatorial sequences exhibiting provably stretched exponential behaviour in their asymptotics for which we have a polynomial time algorithm for their enumeration. We then move on to enumerating the cogrowth sequence of Richard Thompson’s group  $F$ . This is our main interest as it is a long-standing open problem to determine whether  $F$  is amenable [60, 56, 89, 158, 30]. In this case, we improve the algorithm of Haagerup, Haagerup and Ramirez-Solano [128], allowing us to produce 32 terms of the cogrowth sequence - 7 more than were previously known. We then present Guttman’s numerical analysis of these numbers, which strongly suggests that  $F$  is not amenable. Moreover, using results of Pittet and Saloff-Coste [169, 170], we prove that for any real numbers  $\sigma < 1$  and  $\lambda > 1$ , the cogrowth sequence  $(c_n)_{n \in \mathbb{N}_0}$  of Thompson’s group  $F$  satisfies

$$c_n < 16^n \cdot \lambda^{-n^\sigma}$$

for all sufficiently large  $n$ . This implies that if  $F$  is amenable, then the subdominant term in its asymptotics must decrease faster than any stretched exponential. However, numerical evidence strongly suggests that the subdominant term *is* a stretched exponential, providing further evidence that  $F$  is non-amenable. This suggests a potential new path to a proof of non-amenability of  $F$ : if the universality class of the cogrowth sequence can be determined rigorously, it will likely prove non-amenability.

In Chapter 6 we investigate the concept of combinatorial Stieltjes moment sequences. A sequence  $(a_n)_{n \in \mathbb{N}_0}$  is called a Stieltjes moment sequence if there is some probability measure  $\rho$  on  $[0, \infty)$  such that

$$a_n = \int_0^\infty x^n d\rho(x)$$

for each integer  $n \geq 0$ . We prove that the cogrowth sequence for any finitely generated group is a Stieltjes moment sequence, and, more generally, we

prove that this is the case for the counting sequence of returns in any undirected locally-finite graph. We also find empirical evidence that the counting sequences of 1324-avoiding permutations, planar Eulerian orientations and quartic planar Eulerian orientations are all Stieltjes moment sequences. When this can be proved for a sequence  $(a_n)_{n \in \mathbb{N}_0}$ , it immediately produces an increasing sequence of lower bounds  $\mu_n$  for the growth rate  $\mu$ , determined by the terms  $a_0, a_1, \dots, a_n$ . These bounds are stronger than even the ratios of successive terms, which themselves are often difficult to prove to be increasing. In particular, if we were able to prove that the counting sequence for 1324-avoiding permutations was a Stieltjes moment sequence, it would imply a lower bound of 10.302 on its growth rate, which would be an improvement on the bound 10.271 presented in Chapter 3.

In Chapter 7 we investigate a generalisation of Stieltjes moment sequences to sequences of polynomials called Hankel totally positive sequences. It is known that any sequence of polynomials defined by a Thron-type continued fraction (T-fraction) with positive coefficients is Hankel totally positive [188, 166]. We describe a class of objects, called augmented perfect matchings, whose counting function is given by such a T-fraction with infinitely many variables. Hence the associated counting sequence of polynomials is Hankel totally positive. As a special case, we show that this sequence reduces to the Ward polynomials, a sequence of polynomials which counts phylogenetic trees, thereby proving that the Ward polynomials form a Hankel totally positive sequence. We then give a bijective proof that certain weighted phylogenetic trees have the same counting function as augmented perfect matchings.

In Chapter 8 we study the  $O(n)$  loop model on the hexagonal lattice from statistical mechanics. In the case  $n = 0$ , this is equivalent to enumerating self-avoiding walks on the hexagonal lattice. Duminil-Copin and Smirnov proved that the growth constant for self-avoiding walks on the hexagonal lattice is  $\sqrt{2 + \sqrt{2}}$ , using an identity that follows from discrete analogues of the Cauchy-Riemann equations [79]. Smirnov [185] generalised this identity to the  $O(n)$  model with  $n \in [-2, 2]$ , providing an alternative way to predict that the critical point is  $1/\sqrt{2 + \sqrt{2} - n}$ , as conjectured by Nienhuis [162]. We generalise these identities to the off-critical case, allowing us to study the off-critical behaviour of the  $O(n)$  model. In the  $n = 0$  case, this corresponds to the sub-dominant asymptotic behaviour of the counting sequence of self-avoiding walk on the hexagonal lattice. This allows us to relate the winding angle distributions of self-avoiding walks in the half plane and in a wedge with the boundary critical exponents. We find perfect agreement with the conjectured winding angle distribution function predicted by Duplantier and Saleur [81], which relied on non-rigorous Coloumb gas techniques and

conformal invariance.

Problem	Algorithmic complexity	Known terms	Bounds on growth rate	Stieltjes moment sequence
2sip-sortable permutations	Polynomial[3]	> 1000	8.183, 8.346[2]	N
Deque-sortable permutations	Polynomial	> 1000	8.219, 8.352[2]	N
2sis-sortable permutations	$\approx 11.4^n$	20	8.183, 13.374[2]	?(N)
1324-avoiding permutations	$\approx 2^n$ [71]	50[71]	10.271, 13.5	?(Y)
Eulerian orientations	Polynomial	> 1000	$4\pi$	?(Y)
Quartic Eulerian orientations	Polynomial	> 1000	$4\sqrt{3}\pi$	?(Y)
$\mathbb{Z}_2 \wr \mathbb{Z}$ cogrowth	Polynomial	200	9	Y
$\mathbb{Z} \wr_d \mathbb{Z}$ cogrowth	Polynomial	-	$4(d+1)^2$	Y
$B$ cogrowth	Polynomial	128	16	Y
$F$ cogrowth	$\phi^{2n}$ [128]	32	13.269, 16	Y[128]

Table 1.1: A table summarising the current state of some of the enumerative problems considered in this thesis. Original results presented in this thesis are highlighted in blue.

A summary of the enumerative problems considered in this thesis is given in Table 1.1. The second column contains the time complexity of the best known algorithm. The third column contains the number of terms known exactly. The fourth column contains the best proven upper and lower bounds on the growth rate, with only one number when the growth rate is known exactly. The last column contains N if the sequence is not a Stieltjes moment sequence, Y if it is a Stieltjes moment sequence and ? if this is not known. ?(Y) means we conjecture that it is a Stieltjes moment sequence, while ?(N) means we believe that it is not a Stieltjes moment sequence. Original results presented in this thesis are highlighted in blue.

A summary of the predicted asymptotic form for each of these sequences is shown in Table 1.2. For Eulerian orientations and quartic Eulerian orientations, we prove the displayed asymptotic formula in Chapter 4. For the cogrowth sequence of  $\mathbb{Z}_2 \wr \mathbb{Z}$ , the values of  $\mu$ ,  $\kappa$  and  $g$  are known exactly while for  $\mathbb{Z} \wr_d \mathbb{Z}$  the values of  $\mu$ ,  $\kappa$  and  $\delta$  are known exactly. For the Brin-Navas group



Problem	Estimated asymptotic form	Estimated growth rate $\mu$	Estimates of other constants
2sip-sortable permutations	$c\mu^n n^g$	8.281402207	$g \approx -2.473, c \approx 0.080$
Deque-sortable permutations	$c\mu^n n^g$	8.281402207	$c \approx 0.015, g = -3/2$
2sis-sortable permutations	$c\mu^n n^g$	12.4	$g \approx -2.5$
1324-avoiding permutations	$c\mu^n \kappa^{n^\sigma} n^g$	11.60[71]	$\kappa \approx 0.040, \sigma = 1/2, g \approx -1$
Eulerian orientations	$c\mu^n n^g \log(n)^h$	$4\pi$	$c = \pi^2, g = -2, h = -2$
Quartic Eulerian orientations	$c\mu^n n^g \log(n)^h$	$4\sqrt{3}\pi$	$c = 8\pi^2/3, g = -2, h = -2$
$\mathbb{Z}_2 \wr \mathbb{Z}$ cogrowth	$c\mu^n \kappa^{n^\sigma} n^g$	9	$\sigma = 1/3, g = 1/6, \kappa \approx 0.0623, c \approx 0.58$
$\mathbb{Z} \wr_d \mathbb{Z}$ cogrowth	$c\mu^n \kappa^{n^\sigma \log(n)^\delta} n^g$	$4(d+1)^2$	$\sigma = d/(d+2), \delta = 2/(d+2)$
$B$ cogrowth	$c\mu^n \kappa^{n^\sigma \alpha(n)} \beta(n)$	16	$\sigma = 1$
$F$ cogrowth	$c\mu^n \kappa^{n^\sigma \log(n)^\delta} n^g$	$\approx 15$	$\kappa \approx 1/e, \sigma \approx 1/2, \delta \approx 1/2, g \approx -1$

Table 1.2: A table summarising the estimated asymptotic form of some sequences considered in this thesis, based on Guttmann's differential approximant analysis.

$B$ , the values of  $\mu$  and  $\sigma$  are known exactly, but the asymptotic behaviour has not been predicted beyond these terms.  $\alpha(n)$  denotes some function of  $n$  which satisfies  $\alpha(n) \rightarrow 0$  and  $\alpha(n)n^a \rightarrow \infty$  as  $n \rightarrow \infty$  for any  $a > 0$ ; for example  $\alpha(n) = 1/\log(n)$  is a possibility. All other estimates of  $\mu$  and other constants shown in Table 1.2 are non-rigorous predictions.

## Formal power series

In this section we introduce notation and definitions regarding formal power series which will be used throughout this thesis. Let  $A$  be a commutative ring and  $x$  an indeterminate. We denote by  $A[x]$  (resp.  $A[[x]]$ ) the ring of polynomials (resp. formal power series) in  $x$  with coefficients in  $A$ . If  $A$  is a field, then  $A(x)$  denotes the field of rational functions in  $x$ , and  $A((x))$  the

field of Laurent series in  $x$ , that is, series of the form

$$\sum_{n \geq n_0} a_n x^n,$$

with  $n_0 \in \mathbb{Z}$  and  $a_n \in A$ . The coefficient of  $x^n$  in a series  $F(x)$  is denoted by  $[x^n]F(x)$ .

This notation is generalised to polynomials, fractions and series in several indeterminates. For example, if  $F(t, x)$  is the generating function of rooted planar trees where  $t$  counts edges  $x$  counts leaves, then  $F(t, x) \in \mathbb{Q}[x][[t]]$ . For a multivariate series, say  $F(x, y) \in \mathbb{Q}[[x, y]]$ , the notation  $[x^i]F(x, y)$  stands for the series  $F_i(y)$  such that  $F(x, y) = \sum_j F_j(y)x^j$ . It should not be confused with the coefficient of  $x^i y^0$  in  $F(x, y)$ , which we denote by  $[x^i y^0]F(x, y)$ . If  $F(x, x_1, \dots, x_d)$  is a series in the  $x_i$ 's whose coefficients are Laurent series in  $x$ , say

$$F(x, x_1, \dots, x_d) = \sum_{i_1, \dots, i_d} x_1^{i_1} \cdots x_d^{i_d} \sum_{n \geq n_0(i_1, \dots, i_d)} a(n, i_1, \dots, i_d) x^n,$$

then we define the *non-negative part of  $F$  in  $x$*  as the following formal power series in  $x, x_1, \dots, x_d$ :

$$[x^{\geq 0}]F(x, x_1, \dots, x_d) = \sum_{i_1, \dots, i_d} x_1^{i_1} \cdots x_d^{i_d} \sum_{n \geq 0} a(n, i_1, \dots, i_d) x^n.$$

If  $A$  is a field, a power series  $F(x) \in A[[x]]$  is *algebraic* (over  $A(x)$ ) if it satisfies a non-trivial polynomial equation  $P(x, F(x)) = 0$  with coefficients in  $A$ . It is *differentially algebraic* (or *D-algebraic*) if it satisfies a non-trivial polynomial differential equation  $P(x, F(x), F'(x), \dots, F^{(k)}(x)) = 0$  with coefficients in  $A$ . It is *D-finite* if it satisfies a *linear* differential equation with coefficients in  $A(x)$ . For multivariate series, D-finiteness and D-algebraicity require the existence of a differential equation *in each variable*. We refer to [148, 149] for general results on D-finite series, and to [25, Sec. 6.1] for D-algebraic series.

# Chapter 2

## Machine sortable permutations

### Introduction

In his seminal book “The Art of Computer Programming”, Donald Knuth introduced the problem of characterising and enumerating the permutations which can be produced by certain machines. We call these *achievable* permutations. Equivalently one can ask which permutations can be *sorted* by these machines, as these will be exactly the inverses of the achievable permutations. The first non-trivial machine which Knuth considered in this context is a single stack. In this case, starting with the identity permutation  $12 \cdots n$  in the input, two operations are permitted as shown in Fig 2.1:

- The input operation  $I$ , which moves the next element from the input onto the top of the stack.
- The output operation  $O$ , which moves the top element of the stack to the output.

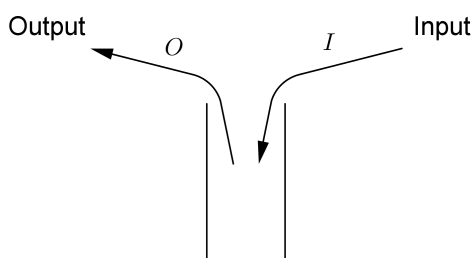


Figure 2.1: The input and output operations  $I$  and  $O$  on a stack.

Then the permutation produced is defined by the order in which the elements are output. In this case, a valid sequence of the operations  $I$  and  $O$  corre-

sponds to a unique Dyck path, with Input operations corresponding to up steps and output operations corresponding to down steps. Moreover, Knuth showed that no two operation sequences produce the same permutation, so the number of permutations of size  $n$  which can be produced by a single stack is equal to the number of Dyck paths of length  $2n$ , which is given by the  $n$ th Catalan number

$$C_n = \frac{1}{n+1} \binom{2n}{n} \sim \frac{1}{\sqrt{\pi}} 4^n n^{-3/2}.$$

Further to this, Knuth characterised the set of achievable permutations, showing that they are exactly the 312-avoiding permutations. Knuth then posed the same problems for each of the following three more complex machines:

- two stacks in parallel (2sip).
- two stacks in series (2sis).
- A double ended queue (deque).

For each of these machines, the set of sortable permutations has been shown to be a permutation class with infinite basis [171, 160]. In other words, the set of sortable permutations cannot be defined as those permutations which avoid some finite set of permutation patterns. For deque-sortable and 2sip-sortable permutations, the (infinitely many) elements of the basis were characterised by Pratt in 1973 [171], but no such characterisation is known for 2sis-sortable permutations. For all three problems a polynomial time algorithm is known for determining whether a permutation is sortable [100, 168]. For  $m \geq 4$ , the problem of determining whether a permutation is sortable by  $m$  stacks in parallel was shown to be NP-complete by Unger [204]. Unger later found a polynomial time algorithm for  $m = 3$  [205]. In both cases Unger relies on the result of Even and Itai [100] that a permutation is sortable by  $m$  stacks in parallel if and only if a corresponding graph is  $m$ -colourable.

For the rest of this chapter we will focus on the problem of enumerating permutations sortable by one of these machines (a deque, 2sis or  $m$ sip for  $m \geq 2$ ). From this enumerative perspective, we find it more convenient to analyse the sequences of operations which sort permutations, rather than analysing the sortable permutations directly. All of these problems remained open until 2015 when Albert and Bousquet-Mélou enumerated permutations sortable by two stacks in parallel [3]. Their solution took the form of a system of functional equations which characterise the counting function  $P(t)$  of these permutations. From an asymptotic perspective, these equations are surprisingly impenetrable, in that we are still unable to determine either

the growth rate or the subexponential behaviour of the sequence. Instead we resort to generating the coefficients of  $P(t)$  exactly and subsequently analysing them. We will outline their solution in more general terms in the next section. In 2017, E. and Guttmann solved the problem for permutations sortable by a deque by demonstrating the following simple algebraic relationship between their counting function  $D(t)$  and that of permutations sortable by two stacks in parallel:

$$2D(t) = 2 + t + 2Pt - 2Pt^2 - t\sqrt{1 - 4P + 4P^2 - 8P^2t + 4P^2t^2 - 4Pt}.$$

We describe our solution in Section 2.3. Using this equation we prove in Subsection 2.3.3 that, subject to three conjectures of Albert and Bousquet-Mélou regarding quarter-plane walks, the growth rates of 2sip-sortable permutations and deque-sortable permutations are equal. We make some further predictions about the asymptotic behaviour of these two sequences, all of which are supported by empirical analysis of the sequences given in Section 2.5.

Finally in Section 2.4 we study the problem of sorting by two stacks in series, which seems to be much harder than the other two problems considered. For this problem we develop an algorithm with which we compute 20 coefficients of the counting function  $S(t)$ . Empirical analysis of these terms is given in Section 2.5.

## Permutations sortable by $m$ stacks in parallel

In this section we will describe part of Albert and Bousquet-Mélou's solution to the problem of enumerating permutations sortable by two stacks in parallel in general terms which apply to any number  $m$  of stacks in parallel. In the case  $m = 1$  we recover the single stack solution, and for  $m = 2$  we get the solution of Albert and Bousquet-Mélou for two stacks in parallel. We will also describe some of the difficulties in solving the problem when  $m > 2$ .

### Notation, operation sequences and perfect matchings

Let the  $m$  stacks be called  $S_1, S_2, \dots, S_m$ . Consider the operations  $I_1, I_2, \dots, I_m$  and  $O_1, O_2, \dots, O_m$ , where  $I_j$  represents input into stack  $S_j$  and  $O_j$  represents output from stack  $j$ . We will call a permutation *msip-achievable* if it can be produced by  $m$  stacks in parallel. Each *msip-achievable* permutation can be encoded as a word  $w$  over the alphabet  $\{I_1, I_2, \dots, I_m, O_1, O_2, \dots, O_m\}$  such that for each integer  $j$ , the number of occurrences of the letter  $I_j$  is equal to the number of occurrences of the letter  $O_j$  and any prefix of  $w$  contains at least as many occurrences of  $I_j$  as  $O_j$ . We will call such a word an *msip-word*.

**Definition 2.1.** An operation sequence is said to *output eagerly* if it contains no consecutive terms  $I_i O_j$  with  $i \neq j$ .

Given an *msip*-achievable permutation  $\pi$  which is produced by an *msip*-word,  $w$ , we can convert  $w$  into an operation sequence which outputs eagerly, without changing the output permutation  $\pi$ , by continually swapping any consecutive letters  $I_i O_j$  with  $i \neq j$ . When no more swaps can be performed, we have an operation sequence which outputs eagerly and produces  $\pi$ . Hence any  $m$ -achievable permutation  $\pi$  can be produced by an *msip* word which outputs eagerly.

**Definition 2.2.** The type of an operation sequence  $w$  is the word over the alphabet  $\{I, O\}$  obtained by deleting all of the subscripts of  $w$ .

**Lemma 2.3.** *If two operation sequences  $u$  and  $v$  produce the same permutation  $\pi$  and both output eagerly, then  $u$  and  $v$  have the same type. Moreover, the  $i$ th operation of  $u$  moves the same item as the  $i$ th operation of  $v$ .*

*Proof.* Let  $u = u_1 u_2 \dots u_{2n}$  and let  $v = v_1 v_2 \dots v_{2n}$ . We will prove the lemma by induction. Assume that for some  $m$ , the words  $u_1 \dots u_m$  and  $v_1 \dots v_m$  have the same type. We will prove that the letters  $u_{m+1}$  and  $v_{m+1}$  also have the same type, and move the same item as each other. After the  $m$ th operation, both  $u$  and  $v$  have the same number of input steps and the same number of output steps as each other, so, since they produce the same sub-permutation  $\pi$ , the items currently in the stacks according to  $u$  must be the same as the items currently in the stacks according to  $v$ , though the arrangement may be different. Since  $u$  and  $v$  output eagerly, they will both output at this point if and only if the next item to be output is currently in one of the stacks. Since this is the same for both  $u$  and  $v$ , the letter  $u_{m+1}$  and  $v_{m+1}$  have the same type. Moreover, if they are both output steps, then the element for each will be the next element to be output, whereas if they are both input steps, the item moved will be the next element in the input. In both cases the item moved is the same for both  $u$  and  $v$ . This completes the induction, so  $u$  and  $v$  have the same type.  $\square$

We will now describe an injection from operation sequences to certain coloured perfect matchings. Recall that a perfect matching of  $[2n]$  is simply a partition of  $[2n]$  into  $n$  pairs. We represent these by a sequence of  $2n$  dots with the  $i$ th and  $j$ th dots joined by an arch if  $\{i, j\}$  is a pair in the perfect matching. If  $i < j$ , we call  $i$  the opener of the arch and  $j$  the closer of the arch. An  $m$ -coloured perfect matching is a perfect matching in which each arch is coloured by one of the  $m$  colours  $c_1, c_2, \dots, c_m$ . We now describe an injection from *msip*-words to  $m$ -coloured perfect matchings:

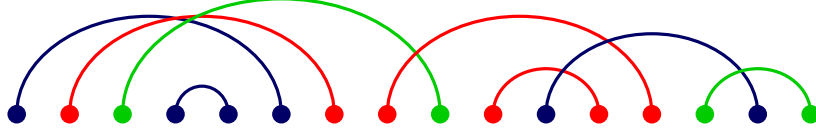


Figure 2.2: The 3-coloured perfect matching corresponding to the operation sequence  $I_1 I_2 I_3 I_1 O_1 O_1 O_2 I_2 O_3 I_2 I_1 O_2 O_2 I_3 O_1 O_3$ . The corresponding 3sip-achievable permutation is 41236578

- Given an *msip*-word  $w$  of length  $2n$ , consider the sorting procedure defined by  $w$  using the identity permutation as input.
- For each item  $i \in [n]$ , if  $i$  is input and output by the  $p$ th and  $q$ th operations respectively, we draw the arch  $(p, q)$  in the perfect matching.
- This arch is given the colour  $c_j$ , where  $S_j$  is the stack which the item is moved to and from. Note that this implies that the  $p$ th and  $q$ th letters of  $w$  are  $I_j$  and  $O_j$ , respectively.

An example of this injection is shown in Figure 2.2.

We will call a perfect matching of  $[2n]$  *properly coloured* if no two crossing arches (that is, arches  $(a, c)$  and  $(b, d)$  with  $a < b < c < d$ ) share the same colour  $c_j$ . Since each  $S_j$  is a stack, the coloured perfect matching corresponding to any *msip*-word must be properly coloured. Moreover, every properly coloured perfect matching has a corresponding operation sequence. We will call a perfect matching *reduced* if whenever an opener is followed immediately by a closer in the perfect matching, they are ends of the same arch. Then a properly coloured perfect matching is reduced if and only if its corresponding operation sequence outputs eagerly.

**Proposition 2.4.** *Two reduced coloured perfect matchings (possibly with different numbers of colours) correspond to the same permutation  $\pi$ , if and only if their underlying (uncoloured) perfect matching is the same.*

*Proof.* Let  $P_1$  and  $P_2$  be the two coloured perfect matchings corresponding to the same permutation  $\pi$ , and let  $\Omega_1$  and  $\Omega_2$  be the corresponding operation sequences. Using Lemma 2.9, we know that  $\Omega_1$  and  $\Omega_2$  have the same type, which means that  $P_1$  and  $P_2$  have the same set of openers  $S_L$  and the same set of closers  $S_R$ . For the openers of  $P_1$ , the corresponding items of the permutation must be ordered from left to right, while for the closers the corresponding items must be in the order  $\pi(1), \pi(2), \dots, \pi(n)$ . The perfect matching is then completely determined, as it must match each opener with

the closer that corresponds to the same item. Hence  $P_2$  must have the same underlying perfect matching.

Now assume that  $P_1$  and  $P_2$  are two proper colourings of the same perfect matching  $P$ , and let  $\pi_1$  and  $\pi_2$  be the corresponding permutations. We will prove that  $\pi_1 = \pi_2$ . The items of  $\pi_1$  corresponding to each arch of  $P_1$  are given by the order of the openers. Then  $\pi_1$  is given by writing the items corresponding to each arch in the order of the closers. As this does not depend on the colouring of the perfect matching, we must have  $\pi_1 = \pi_2$ .  $\square$

**Proposition 2.5.** *The permutations of length  $n$  are in bijection with (un-coloured) reduced perfect matchings of  $[2n]$ , and the number of stacks in parallel required to produce such a permutation  $\pi$  is equal to the number of colours required to properly colour the corresponding reduced perfect matching  $\tau$ .*

*Proof.* Any permutation of length  $n$  can be sorted by  $n$  stacks in parallel, and therefore must have a corresponding reduced, properly  $n$ -coloured perfect matching, and hence a corresponding reduced perfect matching. Conversely, each reduced perfect matching of  $[2n]$  can be properly coloured using  $n$  colours, so it has a corresponding permutation. By the previous Proposition, these correspondences are unique, so they define a bijection between reduced perfect matchings and permutations. Now let  $\pi$  be a permutation and let  $\tau$  be the corresponding reduced perfect matching.  $\pi$  can be produced by  $m$  stacks in parallel if and only if it has a corresponding properly  $m$ -coloured perfect matching, which occurs if and only if  $\tau$  can be properly  $m$ -coloured. Hence the number of stacks in parallel required to produce  $\pi$  is equal to the number of colours required to properly colour  $\tau$ .  $\square$

**Remark:** Given a permutation  $\pi$  of length  $n$  one can construct the corresponding reduced perfect matching by shuffling the sequence of openers  $L_1, L_2, \dots, L_n$  with the sequence of closers  $R_{\pi(1)}, R_{\pi(2)}, \dots, R_{\pi(n)}$ , so that the closers  $R_j$  appear as early as possible in the sequence subject to the condition that  $R_j$  cannot appear before  $L_j$ . Then in the perfect matching, there is an arch between each pair  $R_j, L_j$ . An example of this transformation is shown in Figure 2.3

**Remark:** It follows from this that the question of sortability by  $m$  stacks in parallel is equivalent to the proper  $m$ -colouring question for reduced perfect matchings. For any perfect matching  $\tau$  of  $[2n]$ , there is a reduced perfect matching  $\tau'$  of  $[4n]$  defined by adding a single arch after each opener in  $\tau$ . Moreover, the number of colours required to properly colour  $\tau$  is equal to the equivalent number for  $\tau'$ . Hence the question of sortability by  $m$  stacks in parallel is equivalent to the proper  $m$ -colouring question for perfect matchings. This is equivalent to the  $m$ -colouring question for circle graphs,



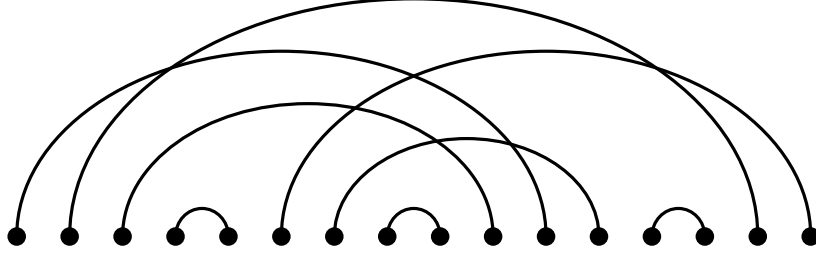


Figure 2.3: The reduced perfect matching corresponding to the permutation 47316825. This perfect matching is properly 3-colourable, so the corresponding permutation is 3sip-sortable.

using a simple correspondence used by Unger [204], so we recover the result of Even and Itai [100] that  $m$ sip-sortability of permutations is equivalent to  $m$ -colourability of corresponding circle graphs. The exact correspondence given by Even and Itai is different and a bit more complicated.

For  $m = 2$ , Albert and Bousquet-Mélou used this construction to enumerate 2sip-sortable permutations [3]. Their solution relates the generating function for 2 colourable reduced perfect matchings (which are equinumerous with 2sip-sortable permutations) with a generating function for 2 coloured perfect matchings. 2 coloured perfect matchings are relatively easy to enumerate as they correspond to quarter-plane loops, using the map sending red openers, blue openers, red closers and blue closers to N,E,S,W steps respectively. Specifically Albert and Bousquet-Mélou proved that the counting function  $P(t)$  of 2sip-sortable permutations is characterised by the following system of equations

$$\begin{aligned} & \left( 1 - ux - uy - \frac{u}{x} - \frac{u}{y} - (a-1)u^2 \left( \frac{x}{y} + \frac{y}{x} \right) \right) \mathcal{Q}(a, u, x, y) \\ &= 1 - \frac{au^2x - u - ux}{y} \mathcal{Q}(a, u, x, 0) - \frac{au^2y - u - uy}{x} \mathcal{Q}(a, u, 0, y) \\ & \mathcal{Q}(a, u^2) = \mathcal{Q}(a, u, 0, 0) \\ & 2P(t) - 1 = \mathcal{Q} \left( \frac{1}{P} - 1, \frac{tP^2}{(1-2P)^2} \right). \end{aligned}$$

In these equations,  $\mathcal{Q}(a, u, x, y)$  is the generating function of square lattice walks confined to the positive quarter-plane where  $a$  counts  $NW$  and  $ES$  corners,  $u$  counts steps and  $x$  and  $y$  count the  $x$  and  $y$  coordinates of the endpoint of the walk.  $\mathcal{Q}(a, u)$  is the generating function of quarter-plane returns to the origin where  $a$  counts  $NW$  and  $ES$  corners and  $u$  counts the half-length.  $\mathcal{Q}(a, u)$  is also the generating function for 2 coloured perfect

matchings where  $u$  counts arches and  $a$  counts occurrences of an opener immediately followed by a closer of a different arch.

The solution for  $m = 2$  stacks in parallel utilises the fact that the number of ways of colouring a 2 colourable arch system is 2 to the power of the number of connected components. For  $m \geq 3$ , it remains easy to enumerate  $m$ -coloured arch systems, using a correspondence with lattice paths in  $m$ -dimensional space confined to the positive cone. The difficulty in enumerating  $m$ sip-sortable permutations occurs because there is no simple formula for the number of proper  $m$ -colourings of an  $m$ -colourable perfect matching. Indeed, Unger showed in [204] that for  $m \geq 4$ , the problem of determining whether a perfect matching is  $m$ -colourable is NP-complete. Unger also showed that there is a polynomial time algorithm in the case  $m = 3$  [205], which perhaps gives some hope of a solution to the enumeration problem of 3sip-sortable permutations.

## Permutations sortable by a double ended queue

In this section we deduce the following equation relating the counting function  $D(t)$  for deque-sortable permutations to the counting function  $P(t)$  for 2sip-sortable permutations:

$$2D(t) = 2 + t + 2Pt - 2Pt^2 - t\sqrt{1 - 4P + 4P^2 - 8P^2t + 4P^2t^2 - 4Pt}.$$

This allows us to characterise  $D(t)$  using a system of functional equations, using the system of functional equations derived by Albert and Bousquet-Mélou [3] which characterise  $P(t)$ .

### Notation and canonical operation sequences

Consider the operations  $I_1, I_2, O_1, O_2$ , where  $I_1$  and  $I_2$  represent input to the top and bottom of the deque, respectively, and  $O_1$  and  $O_2$  represent output from the top and bottom of the deque, respectively, as shown in Fig 2.4. We will call a permutation *deque-achievable* if it can be produced by a deque, and we will call it *2sip-achievable* if it can be produced by two stacks in parallel. Each achievable permutation can be encoded as a word over the alphabet  $\{I_1, I_2, O_1, O_2\}$ , such that the total number of  $I$ 's is the same as the total number of  $O$ 's, and any prefix contains at least as many  $I$ 's as  $O$ 's. We call such a word an *operation sequence*. For example, the permutation 4123 is achievable as it is produced by the operation sequence  $I_1I_1I_1I_2O_2O_2O_2O_2$  (see Fig 2.5). This permutation is not, however, 2sip-achievable.

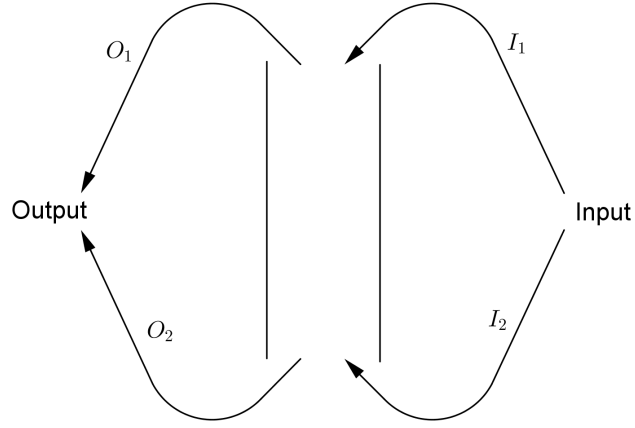


Figure 2.4: The input and output operations  $I_1$ ,  $I_2$ ,  $O_1$  and  $O_2$  on a deque.

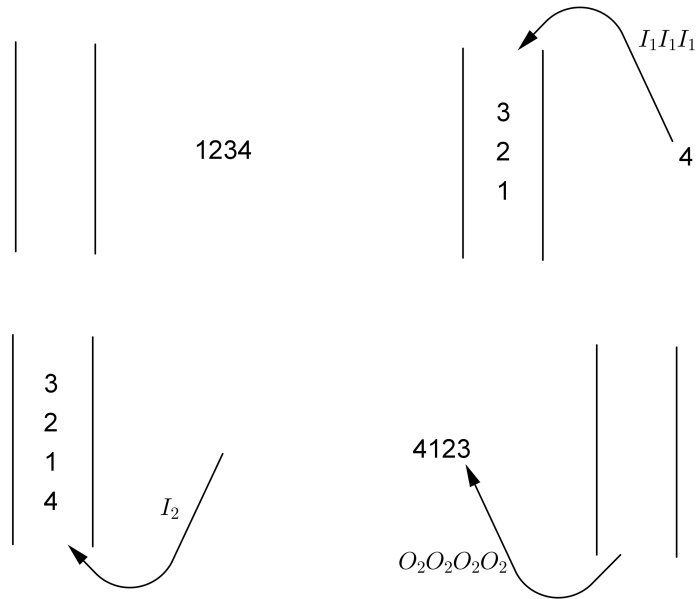


Figure 2.5: The sorting procedure  $I_1I_1I_1I_2O_2O_2O_2O_2$  produces the permutation 4123. This is the shortest permutation which cannot be produced by two stacks in parallel.

We call an operation sequence  $w$  a 2sip word if it has the additional properties that for  $j = 1, 2$  the total number of  $I_j$ 's is equal to the total number of  $O_j$ 's, and each prefix contains at least as many  $I_j$ 's as  $O_j$ 's. In this case, the operation sequence  $w$  corresponds to a 2sip-achievable permutation. Note that if a permutation is 2sip-achievable then it is also deque-achievable, but the converse does not hold.

**Definition 2.6.** We call an operation sequence  $w$  *canonical* if it has the following three properties:

- (outputs eagerly)  $w$  contains no consecutive terms  $I_1O_2$  or  $I_2O_1$ .
- (standard) Any 2sip sub-word of  $w$  begins with  $I_1$ .
- (top happy) Whenever one of the letters  $I_2$  or  $O_2$  appears in  $w$ , the number of preceding  $I$ 's in  $w$  must be at least two greater than the number of preceding  $O$ 's. In other words, these two letters only appear when there are at least two numbers already in the deque.

Note that this definition also defines what is meant by the terms *outputs eagerly*, *standard* and *top happy*.

In Proposition 2.10 we will prove that every deque achievable permutation  $\pi$  is produced by a unique canonical operation sequence. This is analogous to a result in [3], which in our language states that every 2sip achievable permutation  $\pi$  is produced by a unique standard 2sip-word which outputs eagerly.

**Lemma 2.7.** *Any achievable permutation  $\pi$  can be produced by a canonical operation sequence.*

*Proof.* Assign the ordering  $O_1 < O_2 < I_1 < I_2$  to the letters. Since  $\pi$  is achievable, it is produced by some operation sequence. Let  $w$  be the operation sequence which produces  $\pi$  and which is lexicographically minimal. Then we just need to show that  $w$  is canonical.

If  $w$  contains either  $I_1O_2$  or  $I_2O_1$  as a sub-word, then we can switch these letters to produce  $w' < w$ , where  $w'$  also produces  $\pi$ . But this contradicts the minimality of  $w$ , so  $w$  does not contain either  $I_1O_2$  or  $I_2O_1$ . Hence  $w$  outputs eagerly.

Suppose that  $w$  contains a 2sip sub-word  $v$  which begins with  $I_2$ . Then we can form  $v'$  by changing each subscript in  $v$  to the other subscript. Note that  $v'$  begins with  $I_1$ , so  $v' < v$ . Moreover,  $v$  and  $v'$  have the same effect on the state of the machine as each other. Hence, if  $w = avb$ , then the word  $w' = av'b$  produces the same permutation  $\pi$ . But  $w' < w$ , which contradicts

the minimality of  $w$ . So  $w$  does not contain a 2sip sub-word beginning with  $I_2$ . Hence,  $w$  is standard.

Now suppose that  $w$  can be split up as  $w = uv$ , where the number of  $I$ 's in  $u$  is at most one more than the number of  $O$ 's, and where  $v$  begins with either  $I_2$  or  $O_2$ . Then we can form  $v'$  by changing each subscript in  $v$  to the other subscript. Note that  $v'$  begins with  $I_1$  if  $v$  begins with  $I_2$ , and  $v'$  begins with  $O_1$  if  $v$  begins with  $O_2$ , so in either case  $v' < v$ . After  $u$  is applied to the machine, the deque contains at most one element, so  $v'$  and  $v$  have the same effect on the machine, with  $v'$  essentially doing everything on the opposite end of the deque to  $v$ . Hence  $uv'$  produces  $\pi$ . Moreover,  $uv' < uv = w$ , contradicting the minimality of  $w$ . Therefore,  $w$  cannot be split as  $uv$  in this way, so whenever one of the letters  $I_2$  or  $O_2$  appears in  $w$ , the number of preceding  $I$ 's in  $w$  must be at least two greater than the number of preceding  $O$ 's. Hence,  $w$  is top happy.

Therefore,  $w$  is standard, top happy and outputs eagerly. Hence  $w$  is canonical.  $\square$

The proof that this operation sequence is unique is not so straight forward. As in [3] (and Section 2.2), we will consider the type of an operation sequence  $w$ :

**Definition 2.8.** The *type* of an operation sequence  $w$  is the word over the alphabet  $\{I, O\}$  obtained by deleting all of the subscripts of  $w$ .

**Lemma 2.9.** *If two operation sequences  $u$  and  $v$  produce the same permutation  $\pi$  and both output eagerly, then  $u$  and  $v$  have the same type. Moreover, the  $i$ th operation of  $u$  moves the same item as the  $i$ th operation of  $v$ .*

*Proof.* Let  $u = u_1u_2 \dots u_{2n}$  and let  $v = v_1v_2 \dots v_{2n}$ . We will prove the lemma by induction. Assume that for some  $k$ , the words  $u_1 \dots u_k$  and  $v_1 \dots v_k$  have the same type. We will prove that the letters  $u_{k+1}$  and  $v_{k+1}$  also have the same type, and move the same item as each other. After the  $k$ th operation, both  $u$  and  $v$  have the same number of input steps and the same number of output steps as each other, so, since they produce the same sub-permutation  $\pi$ , the items currently in the deque according to  $u$  must be the same as the items currently in the deque according to  $v$ , though the order may be different. Since  $u$  and  $v$  output eagerly, they will both output at this point if and only if the next item to be output is currently in the deque. Since this is the same for both  $u$  and  $v$ , the letter  $u_{k+1}$  and  $v_{k+1}$  have the same type. Moreover, if they are both output steps, then the element for each will be the next element to be output, whereas if they are both input steps, the item moved will be the next element in the input. In both cases the item moved

is the same for both  $u$  and  $v$ . This completes the induction, so  $u$  and  $v$  have the same type.  $\square$

**Proposition 2.10.** *Every deque-achievable permutation  $\pi$  is produced by a unique canonical operation sequence.*

*Proof.* We have already shown that every deque-achievable permutation is produced by some canonical operation sequence, so it remains to show that this operation sequence is unique. Let  $u$  and  $v$  be two canonical operation sequences that produce the permutation  $\pi$ , with  $u \leq v$ . Then we just need to show that  $u = v$ .

Suppose for the sake of contradiction that  $u < v$ . Let  $u = u_1 \dots u_{2n}$  and let  $v = v_1 \dots v_{2n}$ , and let  $k$  be minimal such that  $u_k \neq v_k$ , so  $u_k < v_k$ . Note then that before the  $k$ th move, the position of the deque, input and output is the same according to either  $u$  or  $v$ . By Lemma 2.9,  $u$  and  $v$  have the same type, so the letters  $u_k$  and  $v_k$  have the same type. Therefore,  $u_k, v_k$  are either  $O_1, O_2$ , respectively, or  $I_1, I_2$ , respectively. Since  $v_k$  is  $O_2$  or  $I_2$ , and  $v$  is top happy, there must be at least two items in the deque before this move. Now, by Lemma 2.9, the element moved by  $u_k$  and  $v_k$  must be the same, but  $u_k$  and  $v_k$  act at opposite ends of the deque, so they must both be input operations. Therefore,  $u_k = I_1$  and  $v_k = I_2$ . We will now obtain a contradiction by showing that the operation sequence  $v$  is not standard.

Define the sequences  $k_0, k_1, \dots, k_m$  and  $l_0, l_1, \dots, l_m$  of indices in  $\{k, \dots, 2n\}$  as follows:

- $k_0 = k$
- For each  $i$ , the operations  $v_{k_i}$  and  $v_{l_i}$  move the same item, with  $v_{k_i}$  an input and  $v_{l_i}$  an output move. In particular, this determines  $l_0$ .
- For  $i \geq 0$ , the values  $k_{i+1}$  and  $l_{i+1}$  are chosen inductively so that  $k_i < k_{i+1} < l_i < l_{i+1}$ . Moreover, they are chosen to maximise  $l_{i+1}$ . If no such  $k_{i+1}$  and  $l_{i+1}$  exist, then we set  $m = i$  and terminate. Note that the process must terminate at some point since  $l_1, l_2, \dots$  is a sequence of integers which is increasing and bounded above.

For  $0 \leq i \leq m$ , let  $t_i \in \{1, 2\}$  be the value such that  $v_{k_i} = I_{t_i}$

Let  $D$  be the set of items which are in the deque just before operation  $k$ . Let  $v_j$  be the first operation which outputs an item in  $D$ , and let this item be  $d$ . Then  $d$  must be at one of the ends of the deque just before operation  $k$ . We will prove the following statements for  $0 \leq i \leq m$  in a single induction:

- $k \leq k_i < j$ .

- $u_{k_i} = I_{3-t_i}$ . That is,  $u_{k_i}$  and  $v_{k_i}$  put the element they move at opposite ends of the deque.
- $l_i < j$ .
- $v_{l_i} = O_{t_i}$ . So  $v_{k_i}$  and  $v_{l_i}$  act on the same side of the deque.
- $v_{l_i} = O_{t_i}$ .
- $u_{l_i} = O_{3-t_i}$ . So  $u_{k_i}$  and  $u_{l_i}$  act on the same side of the deque.
- $u_{l_i} = O_{3-t_i}$ .

For the base case, by the definition of  $k$  and  $j$ , we have  $k_0 = k < j$  and  $u_{k_0} = u_k = I_1 = I_{3-t_0}$ .

Now for the inductive step. We will assume that  $k \leq k_i < j$  and  $u_{k_i} = I_{3-t_i}$  and show that  $l_i < j$ ,  $v_{l_i} = O_{t_i}$ ,  $u_{l_i} = O_{3-t_i}$ ,  $k \leq k_{i+1} < j$  and  $u_{k_{i+1}} = I_{3-t_{i+1}}$ . Let  $a_i$  be the item moved by operations  $u_{k_i}, v_{k_i}, u_{l_i}$  and  $v_{l_i}$ . Since  $u_{k_i} = I_{3-t_i}$  and  $v_{k_i} = I_{t_i}$ , we have  $\{u_{k_i}, v_{k_i}\} = \{I_1, I_2\}$ . Therefore, one of  $v_{k_i}$  and  $u_{k_i}$  places  $a_i$  on the end of the deque which  $d$  has to be removed from. So  $a_i$  must be removed before move  $j$ , hence  $l_i < j$ . Since there is an item  $d$  on the deque throughout the whole time that  $a_i$  is on the deque, the input and output operations which move  $a_i$  must act on the same side of the deque. Hence,  $v_{l_i} = O_{t_i}$  and  $u_{l_i} = O_{3-t_i}$ . Now to finish the induction, we just need to show that  $k \leq k_{i+1} < j$  and  $u_{k_{i+1}} = I_{3-t_{i+1}}$ , in the case where  $i < m$ . First,  $k \leq k_i < k_{i+1} < l_i < j$ .

Also, by construction, the output move  $v_{l_i} = O_{t_i}$  for  $a_i$  occurs between the input and output moves for  $a_{i+1}$ , so we must have  $v_{k_{i+1}} = I_{3-t_i}$ , so  $t_{i+1} = 3 - t_i$ . Similarly,  $u_{k_{i+1}} = I_{t_i} = I_{3-t_{i+1}}$ . This completes the induction.

We will now show that  $w = v_k v_{k+1} \dots v_{l_m}$  is a 2sip word. Since  $l_m < j$ , nothing is output from the deque from  $D$ . Now, suppose some item is input but not output by  $w$ . Let the input and output moves of that item be  $v_a$  and  $v_b$ , so  $k < a < l_m < b$  and let  $i$  be maximal such that  $k_i < a$ . If  $i = m$  then  $k_m < a < l_m < b$ , which means that we shouldn't have terminated at  $m$ . If  $i < m$  then  $k_i < a < k_{i+1} < l_i < l_{i+1} < b$ , which contradicts the maximality of  $l_{i+1}$ . Hence every item which is input by  $w$  is also output by  $w$ . Therefore,  $v_k \dots v_{l_m}$  is a 2sip sub-word of  $v$  which begins with  $I_2$ , so  $v$  is not standard. This contradicts the statement that  $v$  is canonical, so we must have  $u = v$ . Therefore, any two canonical operation sequences which produce the same permutation are equal.  $\square$

## Enumeration

In the previous subsection we showed that every deque-achievable permutation is produced by a unique canonical operation sequence, hence the number of deque achievable permutations of length  $n$  is equal to the number of canonical operation sequences of length  $2n$ . Recall that this sequence is given by the generating function  $D(t)$ .

Recall that  $Q(a, u)$  is the generating function for quarter-plane loops, counted by half-length, conjugate to the variable  $u$  and number of NW or ES corners, conjugate to  $a$ . Let  $P(t)$  be the length generating function for permutations which can be produced by two stacks in parallel. We note that this is the same as the generating function for standard 2sip words which output eagerly, counted by half-length. Now using Corollary 9 from [3], we have the following relation between these generating functions:

$$Q\left(\frac{1}{P} - 1, \frac{tP^2}{(1 - 2P)^2}\right) = 2P - 1. \quad (2.1)$$

Now we will define some new generating functions to characterise the length generating function  $D(t)$  of deque-achievable permutations. We will call a non-empty operation sequence *unbreakable* if it contains no 2sip sub-words, apart from possibly itself, and the deque is never empty during the associated procedure. Let  $M(a, u, x)$  be the generating function for top happy, unbreakable operation sequences, counted by half-length, conjugate to  $u$ , number of appearances of a sub-word  $I_1O_2$  or  $I_2O_1$ , conjugate to  $a$ , and number of times when there is only one element in the deque, conjugate to  $x$ .

We define a bi-coloured Dyck path to be a Dyck path where each step is coloured red or blue. A multicoloured peak is an up-step followed by a down-step of the opposite colour. Let  $T(a, u, x)$  be the generating function for bi-coloured Dyck paths, where every step from height 0 or 1 is red, counted by half-length, conjugate to  $u$ , number of multicoloured peaks, conjugate to  $a$ , and number of vertices of the path at height 1, conjugate to  $x$ .

We define a new generating function  $Q_1(a, u, x)$  for quarter-plane loops where every step from the origin is  $N$ , counted by half-length, conjugate to  $u$ , number of NW or ES corners, conjugate to  $a$ , and number of steps to the origin, conjugate to  $x$ . This is related to  $Q$  by  $Q(a, u) = Q_1(a, u, 2)$ . We can also think of  $Q$  and  $Q_1$  as generating functions for 2sip words, by replacing each  $N, E, S, W$  step in the quarter-plane loop with  $I_1, I_2, O_1, O_2$ , respectively. Then  $Q(a, u)$  is the generating function for 2sip words, counted by half-length, conjugate to  $u$  and number of appearances of a sub-word  $I_1O_2$  or  $I_2O_1$ , conjugate to  $a$ . The generating function  $Q_1(a, u, x)$  is similar, except



that it only counts those 2sip words where every letter which appears after an equal number of  $I$ 's and  $O$ 's is an  $I_1$ . The other difference is that  $Q_1$  also counts these words by the number of times there are an equal, non-zero number of  $I$ 's and  $O$ 's, conjugate to  $x$ .

**Lemma 2.11.** *The generating function  $T(a, u, x)$  is given by the equation*

$$T(a, u, x) = \frac{4 + 2xu - 2xau - x + x\sqrt{1 - 12u + 4u^2 - 4au + 4a^2u^2 - 8au^2}}{4 - 2xu - 2xau - x + x\sqrt{1 - 12u + 4u^2 - 4au + 4a^2u^2 - 8au^2}}. \quad (2.2)$$

*Proof.* Let  $T_1(a, u, y)$  be the generating function for bi-coloured Dyck paths, where each step from height 0 is red, counted by half-length, conjugate to  $u$ , multicoloured peaks, conjugate to  $a$ , and number of vertices of the path at height 0, including the end points, conjugate to  $y$ . Let  $T_2(a, u)$  be the generating function for bi-coloured Dyck paths, counted by half-length, conjugate to  $u$ , multicoloured peaks, conjugate to  $a$ . Note that each multicoloured Dyck path  $\omega$  satisfies exactly one of the following:

- $\omega$  is empty,
- $\omega$  is made up of a peak followed by a (possibly empty) multicoloured Dyck path,
- $\omega$  is made up of an up-step, followed by a non-empty multicoloured Dyck path, followed by a down-step, followed by a (possibly empty) multicoloured Dyck path.

The contribution to  $T_2$  from the first case is simply 1.

In the second case, the 4 possible peaks are counted by  $u(2a + 2)$ , so the total contribution to  $T_2$  from this case is  $u(2a + 2)T_2$ .

In the third case, the contribution is simply  $4u(T_2 - 1)T_2$ . The multiplier 4 arises because there are four possibilities for the colours of the up step and down step.  $T_2 - 1$  counts the first (non-empty) path and  $T_2$  counts the second path. The term  $u$  is in the expression because the half length of  $\omega$  is one more than the sum of the half lengths of the shorter paths. Hence we get the equation

$$T_2(a, u) = 4u(T_2 - 1)T_2 + u(2a + 2)T_2 + 1.$$

Solving the quadratic gives

$$T_2(a, u) = \frac{1 + 2u - 2au \pm \sqrt{1 - 12u + 4u^2 - 4au + 4a^2u^2 - 8au^2}}{8u}.$$

Note that since  $T_2(a, 0) = 1$ , we must use the negative square root, so

$$T_2(a, u) = \frac{1 + 2u - 2au - \sqrt{1 - 12u + 4u^2 - 4au + 4a^2u^2 - 8au^2}}{8u}.$$

We now calculate  $T_1(a, u, y)$  in a similar way. Each multicoloured Dyck path  $\omega$ , where each step from height 0 is red, satisfies exactly one of the following:

- $\omega$  is empty,
- $\omega$  is made up of a peak followed by a (possibly empty) multicoloured Dyck path,
- $\omega$  is made up of a red up-step, followed by a non-empty multicoloured Dyck path, followed by a down-step, followed by another (possibly empty) multicoloured Dyck path where each step from height 0 is red.

The contribution to  $T_1$  from the first case is simply  $y$ .

In the second case, the two possible peaks are counted by  $uay + uy$ , so the contribution to  $T_1$  from this case is  $(uay + uy)T_1$ .

In the third case, the half length of  $\omega$  is one more than the sum of the half lengths of the shorter paths and the number of vertices at height 0 is one more than the number of vertices at height 0 in the second path. Hence the contribution from this case is  $2uy(T_2 - 1)T_1$ . Hence, we get the equation

$$T_1(a, u, y) = 2uy(T_2 - 1)T_1 + (uay + uy)T_1 + y.$$

Solving this using the formula for  $T_2$  gives

$$T_1(a, u, y) = \frac{4y}{4 + 2yu - 2yau - y + y\sqrt{1 - 12u + 4u^2 - 4au + 4a^2u^2 - 8au^2}}$$

Finally, we can calculate  $T(a, u, x)$  in a similar way. Each multicoloured Dyck path  $\omega$ , where each step from height 0 or 1 is red, satisfies exactly one of the following:

- $\omega$  is empty
- $\omega$  is made up of a red up-step, followed by a (possibly empty) multicoloured Dyck path, where each step from height 0 is red, followed by a red down-step, followed by another (possibly empty) multicoloured Dyck path where each step from height 0 or 1 is red.

The contribution to  $T$  from the first case is simply 1. In the second case, the half length of  $\omega$  is one more than the sum of the half lengths of the shorter paths, and the number of multicoloured peaks in  $\omega$  is equal to the sum of the numbers of multicoloured peaks in the two shorter paths. The number of vertices at height 1 in the long path is equal to the number of vertices at height 0 in the first short path, plus the number of vertices at height 1 in the second short path. Hence,

$$T(a, u, x) = uT_1(a, u, x) \cdot T(a, u, x) + 1$$

Solving this using the formula for  $T_1$  gives

$$T(a, u, x) = \frac{4 + 2xu - 2xau - x + x\sqrt{1 - 12u + 4u^2 - 4au + 4a^2u^2 - 8au^2}}{4 - 2xu - 2xau - x + x\sqrt{1 - 12u + 4u^2 - 4au + 4a^2u^2 - 8au^2}}$$

as required.  $\square$

**Lemma 2.12.** *The generating function  $Q_1(a, u, x)$  is given by the equation*

$$Q_1(a, u, x) = \frac{2Q}{2Q - xQ + x} \quad (2.3)$$

*Proof.* Let  $U(a, u)$  be the generating function for non-empty quarter-plane loops, which only touch the origin at the start and end, counted by half-length, conjugate to  $u$  and number of  $NW$  or  $ES$  corners, conjugate to  $a$ . Then each loop counted by  $Q$  can be written uniquely as a sequence of loops counted in  $U$ , so we get the equation

$$Q(a, u) = \frac{1}{1 - U(a, u)}.$$

Or, equivalently,

$$U(a, u) = 1 - \frac{1}{Q(a, u)}.$$

Now, by reflecting a loop about the line  $x = y$ , we see that amongst the loops counted by  $U$  with a given half-length and number of  $NW$  or  $ES$  corners, exactly half begin with  $N$ . Hence, each loop counted by  $Q_1$  can be written uniquely as a sequence of loops counted in  $\frac{1}{2}U$ . Moreover, the power of  $x$  in the corresponding monomial in  $Q_1$  is equal to the number of terms in the sequence of loops from  $\frac{1}{2}U$ . Therefore, we have the equation

$$Q_1(a, u, x) = \frac{1}{1 - \frac{1}{2}U(a, u)x}$$

Hence, we can write  $Q_1$  in terms of  $Q$  as desired

$$\begin{aligned} Q_1(a, u, x) &= \frac{1}{1 - \frac{1}{2}(1 - \frac{1}{Q})x} \\ &= \frac{2Q}{2Q - Qx + x} \end{aligned}$$

□

Before we can relate these generating functions to  $M$  and  $D$ , we will need to consider a decomposition of operation sequences, as described in the lemma below. In Lemma 2.15 we will show that this decomposition is unique.

**Lemma 2.13.** *Given a non-empty operation sequence  $w$  it is possible to decompose  $w$  as*

$$w = x_1 w_1 x_2 w_2 \dots x_{2m-1} w_{2m-1} x_{2m} v,$$

where  $m \in \mathbb{Z}_{>0}$ , each  $x_i \in \{I_1, O_1, I_2, O_2\}$ , such that  $x_1 x_2 \dots x_{2m}$  is an unbreakable operation sequence, each  $w_i$  is a (possibly empty) 2sip sub-word, and  $v$  is an operation sequence.

*Proof.* We construct the decomposition as follows: First we decompose  $w$  as  $w = uv$ , so that the first point at which the deque is empty is immediately after  $u$  is applied, in other words,  $u$  is the minimal, non-empty prefix of  $w$  which is also an operation sequence. Note that we may have  $u = w$ . Since the deque is empty before and after  $v$  is applied,  $v$  is an operation sequence.

Now,  $x_1$  must be the first letter of  $u$ . Then, for each  $i$ , we let  $w_i$  be the longest 2sip sub-word of  $u$  which starts immediately after  $x_i$ . Note that this necessarily exists, since  $w_i$  can always be the empty word. Then  $x_{i+1}$  is the letter after  $w_i$ . We continue this until we have decomposed all of  $u$  as  $u = x_1 w_1 \dots w_{j-1} x_j w_j$ . Note that  $u = x_1 w_1 \dots w_j$  and  $w_j$  both contain an equal number of  $I$ 's and  $O$ 's, so  $x_1 w_1 \dots x_j$  also contains an equal number of  $I$ 's and  $O$ 's. But since  $u$  is minimal,  $w_j$  must be empty. Note that by definition, each  $w_i$  is a 2sip sub-word. So we just need to prove that  $x_1 x_2 \dots x_j$  is an unbreakable operation sequence. Since  $x_1 w_1 \dots w_{j-1} x_j$  is an operation sequence, and each  $w_i$  is an operation sequence, it follows that  $x_1 x_2 \dots x_j$  is an operation sequence. Also, if  $x_1 x_2 \dots x_i$  is an operation sequence for any  $i < j$ , then  $x_1 w_1 \dots x_i$  is also an operation sequence, which contradicts the minimality of  $u$ . Now we just need to show that  $x_1 \dots x_j$  contains no 2sip sub-words, other than itself and the empty word. Suppose that it does contain some 2sip sub-word  $x_a \dots x_i$ , with  $i \geq a$ . If  $a > 1$  then  $w_{a-1} x_a w_a x_{a+1} \dots x_i$  is also a 2sip sub-word, which contradicts the maximality of the length of

$w_{a-1}$ . If  $a = 1$  and  $i < j$ , then  $x_a x_{a+1} \dots x_i$  is not an operation sequence, so it is certainly not a 2sip sub-word. Hence  $a = 1$  and  $i = j$ , so  $x_1 x_2 \dots x_j$  contains no 2sip sub-words other than possibly itself. Therefore  $x_1 x_2 \dots x_j$  is unbreakable.  $\square$

**Lemma 2.14.** *Let  $w$  be a non-empty operation sequence with the decomposition*

$$w = x_1 w_1 x_2 w_2 \dots x_{2m-1} w_{2m-1} x_{2m} v,$$

*where  $m \in \mathbb{Z}_{>0}$ , each  $x_i \in \{I_1, O_1, I_2, O_2\}$ , such that  $x_1 x_2 \dots x_{2m}$  is an unbreakable operation sequence, each  $w_i$  is a (possibly empty) 2sip sub-word, and  $v$  is an operation sequence. Then any 2sip sub-word of  $w$ , which is not a prefix, is contained in one of the words  $w_i$  or  $v$ .*

*Proof.* Let  $u = x_1 w_1 x_2 w_2 \dots x_{2m-1} w_{2m-1} x_{2m}$ . Suppose for the sake of contradiction that some 2sip sub-word  $w'$  of  $w$  is not contained in any  $w_i$  or  $v$ , and  $w'$  is not a prefix of  $w$ .

First we consider the case where  $w'$  does not intersect with  $v$ . Then let  $w' = s x_i w_i x_{i+1} w_{i+1} \dots x_j t$  be a longer 2sip sub-word of  $u$ , where  $t$  is a prefix of  $w_j$  and  $s$  is a suffix of  $w_{i-1}$ . Since  $t$  is a suffix of an operation sequence  $w'$  and a prefix of another operation sequence, it must be an operation sequence. Moreover, since  $t$  is a prefix of a 2sip word,  $t$  is also a 2sip word. Similarly,  $s$  is a 2sip word. Hence,  $s, w_i, w_{i+1}, \dots, w_{j-1}, t$  are all 2sip words, as is  $w' = s w_i x_{i+1} w_{i+1} \dots w_{j-1} x_j t$ , so  $x_{i+1} x_{i+2} \dots x_j$  is also a 2sip word. But this is a contradiction, since  $x_1 x_2 \dots x_{2m}$  is unbreakable.

Now we consider the case where  $w'$  intersects with  $v$ . Then we can write  $w' = u' v'$ , where  $u'$  is a suffix of  $u$  and  $v'$  is a prefix of  $v$ . Since  $w'$  is not a prefix of  $w$ , the word  $u'$  is also not a prefix. Since  $w'$  is not contained in  $v$ , the word  $u'$  is non-empty. Since  $u'$  is a prefix of an operation sequence and a suffix of an operation sequence, it is also an operation sequence. But then  $u'$  satisfies the properties of  $w'$  in the first case, a contradiction.  $\square$

**Lemma 2.15.** *Given a non-empty operation sequence  $w$  there is a unique way to decompose  $w$  as*

$$w = x_1 w_1 x_2 w_2 \dots x_{2m-1} w_{2m-1} x_{2m} v,$$

*where  $m \in \mathbb{Z}_{>0}$ , each  $x_i \in \{I_1, O_1, I_2, O_2\}$ , such that  $x_1 x_2 \dots x_{2m}$  is an unbreakable operation sequence, each  $w_i$  is a (possibly empty) 2sip sub-word, and  $v$  is an operation sequence.*

*Proof.* Let  $w = x_1 w_1 x_2 w_2 \dots x_{2m-1} w_{2m-1} x_{2m} v$  be one such decomposition of  $w$ . We just need to show that this is the same as the decomposition constructed in Lemma 2.13.

First we will show that  $u = x_1w_1x_2w_2 \dots x_{2m-1}w_{2m-1}x_{2m}$  is the shortest operation sub-sequence starting from the start of  $w$ . Since each  $w_i$  is an operation sequence, and  $x_1 \dots x_{2m}$  is an operation sequence,  $x_1w_1x_2w_2 \dots x_{2m-1}w_{2m-1}x_{2m}$  is an operation sequence. Let  $u'$  be a non-empty prefix of  $u$  such that  $u' \neq u$ . Then  $u' = x_1w_1 \dots x_{i-1}w_{i-1}x_it$  for some  $i < 2m$ , where  $t$  is a prefix of  $w_i$ . Then, since  $x_1 \dots x_{2m}$  is unbreakable, there are strictly more  $I$ 's than  $O$ 's in  $x_1 \dots x_i$ . Also, each  $w_k$  contains an equal number of  $I$ 's and  $O$ 's and  $t$  contains at least as many  $I$ 's as  $O$ 's. Hence  $u'$  contains more  $I$ 's than  $O$ 's, so it is not an operation sequence. Therefore,  $u$  is the shortest non-empty prefix of  $w$ , which is also an operation sequence.

Finally, by the previous lemma, every 2sip sub-word of  $w$ , is either a prefix of  $w$  or is contained in one of the words  $w_i$  or  $v$ . Hence, each  $w_i$  is the 2sip sub-word of maximal length with that starting point. Therefore, this decomposition is the same as the decomposition constructed in Lemma 2.13  $\square$

**Lemma 2.16.** *Let  $w$  be a non-empty operation sequence, with decomposition  $w = x_1w_1 \dots w_{2m-1}x_{2m}v$ .  $w$  is canonical if and only if the following conditions hold:*

- *The (unbreakable) operation sequence  $s = x_1x_2 \dots x_{2m}$  is top happy,*
- *Each 2sip word  $w_i$  is standard and outputs eagerly,*
- *$v$  is canonical,*
- *If some  $x_ix_{i+1}$  is either  $I_1O_2$  or  $I_2O_1$ , then  $w_i$  is non-empty.*

*Proof.* If  $w$  is canonical, then  $w$  is top happy, so  $s$  is top happy and  $v$  is top happy. Since  $w$  is standard and outputs eagerly, any sub-operation sequence of  $w$  is also standard and outputs eagerly; in particular this includes each  $w_i$ , as well as  $v$ . Hence  $v$  is canonical. Finally the fourth condition follows immediately from the fact that  $w$  outputs eagerly.

Now we will assume the four conditions and prove that  $w$  is canonical. A sub-word of the form  $IO$  can only appear in  $w$  inside one of the sub-words  $w_i$  or  $v$  or as  $x_ix_{i+1}$ , where  $w_i$  is empty. The conditions clearly make it impossible for such a sub-word to be either  $I_1O_2$  or  $I_2O_1$ , so  $w$  outputs eagerly. Now we consider  $w$  to be a bi-coloured Dyck path. Since  $s$  and  $v$  are top happy, every step from height 0 or 1 in  $w$  which comes from  $s$  or  $v$  is red. Any step from height 0 or 1 which comes from some  $w_i$ , must be at height 1 in  $w$  and height 0 in  $w_i$ . Then this step must be red since  $w_i$  is standard. Finally, by Lemma 2.14, any 2sip sub-word of  $w$  is either a prefix of  $w$ , contained in  $v$  or contained in one of the words  $w_i$ . Hence, since each

$w_i$  and  $v$  are standard, and  $w$  begins with  $I_1$ , the word  $w$  is also standard. Therefore,  $w$  is canonical.  $\square$

Now recall that the generating function  $M(a, u, x)$  is the generating function for top happy, unbreakable operation sequences, counted by half-length, conjugate to  $u$ , number of appearances of a sub-word  $I_1O_2$  or  $I_2O_1$ , conjugate to  $a$ , and number of times when there is only one element in the deque, conjugate to  $x$ .

**Lemma 2.17.** *The generating function  $M$  satisfies the equation*

$$T(a, u, x) = M\left(1 + \frac{a-1}{Q}, uQ^2, \frac{Q_1}{Q}x\right) \frac{T}{Q} + 1 \quad (2.4)$$

*Proof.* We first note that given an operation sequence  $w$ , there is a corresponding bi-coloured Dyck path, formed by replacing each  $I_1$  with a red up-step, each  $I_2$  with a blue up-step, each  $O_1$  with a red down-step, and each  $O_2$  with a blue down-step. Then the condition that the steps from height 0 or 1 in the Dyck paths counted by  $T$  are all red is equivalent to the condition that the corresponding operation sequence is top happy. Hence, we can consider  $T(a, u, x)$  to be the generating function for top happy operation sequences, counted by half-length, conjugate to  $u$ , number of consecutive steps  $I_1O_2$  or  $I_2O_1$ , conjugate to  $a$ , and number of times in the procedure that the deque contains exactly one element, conjugate to  $x$ .

For each non-empty operation sequence  $w$  counted by  $T(a, u, x)$  we consider the decomposition  $w = x_1w_1x_2w_2 \dots x_{2m-1}w_{2m-1}x_{2m}v$  described in Lemma 2.15. In particular, we consider the contribution to  $T$  from all words  $w$  with a given unbreakable operation sequence  $s = x_1x_2 \dots x_{2m}$ . Since  $w$  is top happy,  $s$  is also top happy, so we will only consider the top happy, unbreakable operation sequences  $s$ , which are exactly the operation sequences counted by  $M$ . Let  $q, r$  be the number of consecutive  $I_1O_2$  or  $I_2O_1$  steps and number of times in the procedure given by  $s$  that the deque contains exactly one element respectively. So the contribution of  $s$  to  $M(a, u, x)$  is  $a^qu^mx^r$ .

Now we calculate the contribution to  $T$  from all the words of the form  $w$ , with  $s = x_1x_2 \dots x_{2m}$  fixed. Call this the contribution of  $s$  to  $T$ . Each  $w_i$  which begins (and ends) at height 1 in the bi-coloured Dyck path corresponding to  $w$  can be any 2sip word whose steps from height 0 in its bi-coloured Dyck path are all red. Hence the possible words  $w_i$  are exactly those counted by  $Q_1$ . Each  $w_i$  which begins and ends with the deque containing more than one element can be any 2sip word, and these are enumerated by  $Q$ . Also, the possible words  $v$  are exactly those counted by  $T$ . Since the half-length of  $w$

is  $m$  plus the sum of the half-lengths of these sub-words, the contribution of  $s$  to  $T(1, u, 1)$  is

$$u^m Q_1(1, u, 1)^r Q(1, u)^{2m-r-1} T(1, u, 1).$$

Now, each vertex at height 1 in the bi-coloured Dyck path corresponding to  $w$  occurs in one of the words  $w_i$  counted by  $Q_1$ , or in  $v$ . The number of these vertices in any word  $w_i$  counted by  $Q_1$ , is equal to the number of vertices in the bi-coloured Dyck path of  $w_i$  at height 0, which is 1 more than the power of  $x$  in the contribution of  $w_i$  to  $Q_1(a, u, x)$ . Since all other vertices at height 1 appear in  $v$ , the contribution of  $s$  to  $T(1, u, x)$  is equal to

$$u^m (xQ_1(1, u, x))^r Q(1, u)^{2m-r-1} T(1, u, x).$$

Now we just need to consider the number of sub-words  $I_1O_2$  and  $I_2O_1$  in  $w$ . If we only consider the sub-sequences which occur in one of the words  $w_i$  or  $v$ , we would get the contribution

$$u^m (xQ_1(a, u, x))^r Q(a, u)^{2m-r-1} T(a, u, x).$$

The only other situation where one of these sub-sequences occurs is when one of the words  $w_i$  is empty, and the surrounding letters  $x_i x_{i+1}$  form one of these sub-words. Note that this cannot happen at height 1, since  $w$  is top happy, so it only occurs in the case where  $w_i$  is counted by  $Q$ . This case is counted as 1 in  $Q(a, u)$  in the equation above, but it should be counted as  $a$ . Moreover, this is relevant to exactly  $q$  of the words  $w_i$ , exactly those where  $x_i x_{i+1}$  is either the word  $I_1O_2$  or  $I_2O_1$ . Hence, the contribution of  $s$  to  $T(a, u, x)$  is

$$u^m (xQ_1(a, u, x))^r Q(a, u)^{2m-r-q-1} (Q(a, u) + a - 1)^q T(a, u, x).$$

Therefore, any top happy, unbreakable operation sequence  $s$  which contributes  $a^q u^m x^r$  to  $M(a, u, x)$ , contributes

$$\left(1 + \frac{a-1}{Q}\right)^q (uQ^2)^m \left(\frac{Q_1}{Q}x\right)^r \frac{T}{Q}$$

to  $T(a, u, x)$ , and this accounts for all of  $T(a, u, x)$  except for the 1 coming from the empty word. Hence,

$$T(a, u, x) = M\left(1 + \frac{a-1}{Q}, uQ^2, \frac{Q_1}{Q}x\right) \frac{T}{Q} + 1.$$

□



**Lemma 2.18.** *The generating function  $D$  satisfies the equation*

$$D(t) = M\left(1 - \frac{1}{P}, tP^2, 1\right) \frac{D}{P} + 1. \quad (2.5)$$

*Proof.* Using Proposition 2.10, we know that  $D(t)$  is the generating function for canonical operation sequences, counted by half-length. Using Lemma 2.16, we see that  $D(t)$  is the generating function for words  $w$  which decompose as  $w = x_1w_1 \dots w_{2m-1}x_{2m}v$ , where

- The (unbreakable) operation sequence  $s = x_1x_2 \dots x_{2m}$  is top happy,
- Each 2sip word  $w_i$  is standard and outputs eagerly,
- $v$  is canonical,
- If some  $x_ix_{i+1}$  is either  $I_1O_2$  or  $I_2O_1$ , then  $w_i$  is non-empty.

As in the proof of the previous lemma, we will consider the contribution of any given top happy, unbreakable operation sequence  $s = x_1x_2 \dots x_{2m}$  to  $D$ , assuming that  $s$  is counted by the monomial  $a^qu^mx^r$  in  $M(a, u, x)$ . For each  $i$ , the word  $w_i$  can be any standard 2sip word which outputs eagerly, except that it can't be empty if  $x_ix_{i+1}$  is  $I_1O_2$  or  $I_2O_1$  and there are  $q$  such values of  $i$ . Also,  $v$  can be any canonical operation sequence. Recall that standard 2sip words which output eagerly are counted by  $P$ . Therefore, the contribution of  $s$  to  $D(t)$  is

$$t^m P^{2m-q-1} (P-1)^q D = \left(1 - \frac{1}{P}\right)^q (tP^2)^m \frac{D}{P}.$$

Hence,

$$D(t) = M\left(1 - \frac{1}{P}, tP^2, 1\right) \frac{D}{P} + 1.$$

□

**Theorem 2.19.** *Let  $D(t)$  be the length generating function for permutations which are sortable by a deque, and let  $P(t)$  be the length generating function for permutations which are sortable by two stacks in parallel. Then  $D$  and  $P$  satisfy the following two (equivalent) equations:*

$$P(t) = \frac{(D-1)(D-t-1)}{2t(D-1-Dt)} \quad (2.6)$$

and

$$2D(t) = 2 + t + 2Pt - 2Pt^2 - t\sqrt{1 - 4P + 4P^2 - 8P^2t + 4P^2t^2 - 4Pt}. \quad (2.7)$$

*Proof.* First we substitute (2.3) into (2.4), to remove  $Q_1$ :

$$T(a, u, x) = M \left( 1 + \frac{a-1}{Q}, uQ^2, \frac{2x}{2Q - xQ + x} \right) \frac{T}{Q} + 1.$$

Now we combine this with (2.1) to get

$$T \left( \frac{1}{P} - 1, \frac{tP^2}{(1-2P)^2}, x \right) = M \left( 1 - \frac{1}{P}, tP^2, \frac{x}{2P-1-Px+x} \right) \frac{T}{2P-1} + 1.$$

Therefore,

$$T \left( \frac{1}{P} - 1, \frac{tP^2}{(1-2P)^2}, 2 - \frac{1}{P} \right) = M \left( 1 - \frac{1}{P}, tP^2, 1 \right) \frac{T}{2P-1} + 1.$$

Using (2.5), we can write this  $M$  in terms of  $D$  and  $P$ . So

$$T \left( \frac{1}{P} - 1, \frac{tP^2}{(1-2P)^2}, 2 - \frac{1}{P} \right) = \frac{(D-1)P}{D} \cdot \frac{T}{2P-1} + 1.$$

Solving for  $T$  gives the relation

$$T \left( \frac{1}{P} - 1, \frac{tP^2}{(1-2P)^2}, 2 - \frac{1}{P} \right) = \frac{(2P-1)D}{DP + P - D}. \quad (2.8)$$

Finally, using the formula (2.2) for  $T$  and rearranging gives the desired result.  $\square$

## Analysis

The main purpose of this subsection is to reduce the problem of showing that the generating functions  $P$  and  $D$  have the same radius of convergence to a few conjectures about the generating function  $Q(a, u)$  for quarter-plane loops. We will also give some evidence for the stronger conjecture below.

**Conjecture 2.20.** *For  $n \in \mathbb{Z}_{\geq 0}$ , let  $p_n$  be the number of permutations of size  $n$  which are sortable by two stacks in parallel and let  $d_n$  be the number of permutations of size  $n$  which are sortable by a deque. We conjecture based on Theorem 2.19 that  $p_n \sim \text{const} \cdot \mu^n \cdot n^\gamma$  and  $d_n \sim \text{const} \cdot \mu^n \cdot n^{-3/2}$  for some constants  $\mu$  and  $\gamma$ .*

We first list three Conjectures of Albert and Bouquet-Mélou which are needed to show the above conjectures. The following are conjectures 10, 11 and 12 in [3], respectively.

**Conjecture 2.21.** *The series  $Q(a, u)$  is  $(a + 1)$ -positive. That is,  $Q$  takes the form*

$$Q(a, u) = \sum_{n \geq 0} u^n P_n(a + 1),$$

*where each polynomial  $P_n$  has positive coefficients.*

**Conjecture 2.22.** *the radius of convergence  $\rho_Q(a)$  of  $Q(a, \cdot)$  is given by*

$$\rho_Q(a) = \begin{cases} \frac{1}{(2 + \sqrt{2 + 2a})^2}, & \text{if } a \geq -1/2, \\ \frac{-a}{2(a - 1)^2}, & \text{if } a \in [-1, -1/2]. \end{cases}$$

**Conjecture 2.23.** *The series  $Q_u(a, u) = \frac{\partial Q}{\partial u}$  is convergent at  $u = \rho_Q(a)$  for  $a \geq -1/3$ .*

Let  $t_c$  denote the radius of convergence of the generating function  $P(t)$ . Theorem 23 in [3] states that assuming the two Conjectures 2.21 and 2.23 are both true, the inequality

$$\frac{t}{(2 - \frac{1}{P(t)})^2} \leq \rho_Q(1/P(t) - 1)$$

holds for  $0 \leq t \leq t_c$ , with equality if and only if  $t = t_c$ , and

$$P(t_c) \leq \frac{3}{2}.$$

Moreover, the following corollary in [3] states that assuming the last three conjectures are all true, the following equation holds:

$$\sqrt{2P(t_c)} = 1 + \sqrt{2t_c P(t_c)}. \quad (2.9)$$

Note that in [3]  $S$  is used instead of  $P$  to denote the generating function for permutations sortable by two stacks in parallel.

**Theorem 2.24.** *Assuming the conjectures 2.21, 2.22 and 2.23, the number of permutations of size  $n$  sortable by two stacks in parallel and the number of permutations sortable by a deque of size  $n$  have the same exponential growth rate.*

*Proof.* It suffices to prove that the generating functions  $P(t)$  and  $D(t)$  have the same radius of convergence  $t_c$ . Since every permutation which is sortable by two stacks in parallel is also sortable by a deque, the coefficients of  $D(t)$

are no smaller than the coefficients of  $P(t)$ . Hence the radius of convergence  $t_D$  of  $D(t)$  satisfies  $t_D \leq t_c$ . Therefore, it suffices to prove that  $D(t)$  is convergent for  $t \in [0, t_c)$ .

Since  $P(t)$  is convergent for  $t \in [0, t_c)$ , the series

$$1 - 4P(t) + 4P(t)^2 - 8P(t)^2t + 4P(t)^2t^2 - 4P(t)t$$

is also convergent in this region. Since  $P(t)$  has positive coefficients, it is increasing on the interval  $[0, t_c]$ . Hence,

$$1 = P(0) \leq P(t) \leq P(t_c) \leq \frac{3}{2}$$

inside this interval. Therefore, since we are assuming Conjecture 2.22, we have the inequality

$$\frac{t}{(2 - \frac{1}{P(t)})^2} \leq \rho_Q(1/P(t) - 1) = \frac{1}{(2 + \sqrt{2/P(t)})^2},$$

where equality holds if and only if  $t = t_c$ . Taking square roots on both sides and rearranging, noting that  $2(2 - 1/P(t)) = (2 + \sqrt{2/P(t)})(2 - \sqrt{2/P(t)})$ , we get the inequality

$$\sqrt{2P(t)} \geq 1 + \sqrt{2tP(t)},$$

with equality if and only if  $t = t_c$ . Now we can remove the square roots as follows to get the inequality

$$\begin{aligned} 0 &\leq (\sqrt{2P} - 1 - \sqrt{2tP})(\sqrt{2P} + 1 - \sqrt{2tP})(\sqrt{2P} - 1 + \sqrt{2tP})(\sqrt{2P} + 1 + \sqrt{2tP}) \\ &= 1 - 4P(t) + 4P(t)^2 - 8P(t)^2t + 4P(t)^2t^2 - 4P(t)t, \end{aligned}$$

with equality if and only if  $t = t_c$ . Since the series

$$1 - 4P(t) + 4P(t)^2 - 8P(t)^2t + 4P(t)^2t^2 - 4P(t)t$$

is convergent and positive for  $t \in [0, t_c)$ , the series

$$\sqrt{1 - 4P(t) + 4P(t)^2 - 8P(t)^2t + 4P(t)^2t^2 - 4P(t)t}$$

is also convergent in this domain. Hence, the series

$$2D(t) = 2 + t + 2Pt - 2Pt^2 - t\sqrt{1 - 4P + 4P^2 - 8P^2t + 4P^2t^2 - 4Pt}$$

is also convergent in this domain. □

For the following analysis, we will assume that

$$P(t) = c_0 + c_1(1 - t/t_c) + c_2(1 - t/t_c)^\alpha + o((1 - t/t_c)^\alpha), \quad (2.10)$$

for some constants  $c_0$ ,  $c_1$ ,  $\alpha$  and  $c_2$ , with  $c_2 \neq 0$ . In Section 2.5 we will present numerical evidence that  $\alpha \approx 1.473$ , but for the following theorem we will assume only that  $2 > \alpha > 1$ .

**Theorem 2.25.** *Assuming that  $P$  takes the form given in (2.10), and assuming the three conjectures 2.21, 2.22 and 2.23, we have*

$$D(t) = D(t_c) + b_1(1 - t/t_c)^{1/2} + o((1 - t/t_c)^{1/2}),$$

where  $D(t_c)$  and  $b_1$  are given by the equations

$$D(t_c) = 2 + t_c + 2c_0t_c - 2c_0t_c^2$$

and

$$b_1 = -t_c \sqrt{(2c_0)^{3/2}t_c - 2c_1\sqrt{t_c}}.$$

*Proof.* First, using (2.7), we can rewrite (2.9) as

$$\sqrt{2c_0} = 1 + \sqrt{2t_cc_0}.$$

It follows that

$$\begin{aligned} 0 &= (1 + \sqrt{2t_cc_0} - \sqrt{2c_0})(1 + \sqrt{2t_cc_0} + \sqrt{2c_0})(1 - \sqrt{2t_cc_0} - \sqrt{2c_0})(1 - \sqrt{2t_cc_0} + \sqrt{2c_0}) \\ &= 1 - 4c_0 + 4c_0^2 - 8c_0^2t_c + 4c_0^2t_c^2 - 4c_0t_c. \end{aligned}$$

Now, recall that

$$2D(t) = 2 + t + 2Pt - 2Pt^2 - t\sqrt{1 - 4P + 4P^2 - 8P^2t + 4P^2t^2 - 4Pt}.$$

Using (2.10) we can expand the expression under the square root as a power series in  $(1 - t/t_c)$ :

$$1 - 4P + 4P^2 - 8P^2t + 4P^2t^2 - 4Pt = q_1(1 - t/t_c) + q_2(1 - t/t_c)^\alpha + o((1 - t/t_c)^\alpha),$$

where  $q_1$  and  $q_2$  are constants defined by

$$q_1 = -4c_1 + 8c_0c_1 - 16c_0c_1t_c + 8c_0c_1t_c^2 - 4c_1t_c + 8c_0^2t_c + 4c_0t_c - 8c_0^2t_c^2$$

and

$$q_2 = -4c_2 + 8c_0c_2 - 16c_0c_2t_c + 8c_0c_2t_c^2 - 4c_2t_c.$$

Since this expression is non-negative for  $t \in [0, t_c]$ , we must have  $q_1 \geq 0$ .

Taking the square root of this, we get

$$\begin{aligned} & \sqrt{1 - 4P + 4P^2 - 8P^2t + 4P^2t^2 - 4Pt} \\ &= (1 - t/t_c)^{1/2} \sqrt{q_1 + q_2(1 - t/t_c)^{\alpha-1} + o((1 - t/t_c)^{\alpha-1})} \\ &= \sqrt{q_1}(1 - t/t_c)^{1/2} + o((1 - t/t_c)^{1/2}). \end{aligned}$$

Finally, we can use this to determine the asymptotics of  $D$

$$\begin{aligned} 2D(t) &= 2 + t + 2Pt - 2Pt^2 - t\sqrt{1 - 4P + 4P^2 - 8P^2t + 4P^2t^2 - 4Pt} \\ &= (2 + t_c + 2c_0t_c - 2c_0t_c^2) - t_c\sqrt{q_1}(1 - t/t_c)^{1/2} + o((1 - t/t_c)^{1/2}). \end{aligned}$$

Simplifying this using the equation  $\sqrt{2c_0} = 1 + \sqrt{2t_cc_0}$  yields the desired expressions for  $D(t_c)$  and  $b_1$ .  $\square$

**Theorem 2.26.** *With the same assumptions as in the previous theorem, and the additional assumptions that  $b_1 > 0$  and  $\alpha < 3/2$ , we have the expansion*

$$D(t) = D(t_c) + b_1(1 - t/t_c)^{1/2} + b_2(1 - t/t_c)^{\alpha-1/2} + o((1 - t/t_c)^{\alpha-1/2}),$$

where  $b_2$  is given by the equation

$$b_2 = \frac{-t_c^{3/2}c_2}{\sqrt{(2c_0)^{3/2}t_c - 2c_1\sqrt{t_c}}}.$$

*Proof.* From the proof of the previous theorem we have

$$\begin{aligned} & \sqrt{1 - 4P + 4P^2 - 8P^2t + 4P^2t^2 - 4Pt} \\ &= (1 - t/t_c)^{1/2} \sqrt{q_1 + q_2(1 - t/t_c)^{\alpha-1} + o((1 - t/t_c)^{\alpha-1})} \end{aligned}$$

Since  $q_1 > 0$ , we can expand this as a power series in  $(1 - t/t_c)$  as follows

$$\begin{aligned} & (1 - t/t_c)^{1/2} \sqrt{q_1 + q_2(1 - t/t_c)^{\alpha-1} + o((1 - t/t_c)^{\alpha-1})} \\ &= (1 - t/t_c)^{1/2} \left( \sqrt{q_1} + \frac{q_2}{2\sqrt{q_1}}(1 - t/t_c)^{\alpha-1} + o((1 - t/t_c)^{\alpha-1}) \right) \\ &= \sqrt{q_1}(1 - t/t_c)^{1/2} + \frac{q_2}{2\sqrt{q_1}}(1 - t/t_c)^{\alpha-1/2} + o((1 - t/t_c)^{\alpha-1/2}). \end{aligned}$$

Therefore, we have

$$\begin{aligned} 2D(t) &= 2 + t + 2Pt - 2Pt^2 - t\sqrt{1 - 4P + 4P^2 - 8P^2t + 4P^2t^2 - 4Pt} \\ &= (2 + t_c + 2c_0t_c - 2c_0t_c^2) - t_c\sqrt{q_1}(1 - t/t_c)^{1/2} \\ &\quad + \frac{t_cq_2}{2\sqrt{q_1}}(1 - t/t_c)^{\alpha-1/2} + o((1 - t/t_c)^{\alpha-1/2}). \end{aligned}$$

Simplifying this using the equation  $\sqrt{2c_0} = 1 + \sqrt{2t_cc_0}$  yields the desired expression for  $b_2$ .  $\square$

## Algorithms for 2sip and deque-sortable permutations

In this section we describe the algorithm we used to generate  $N = 1336$  coefficients of the 2sip series  $P(t)$  and deque series  $D(t)$ .

Recall the equation of Albert and Bousquet-Mélou [3] relating the generating  $P(t)$  of 2sip-sortable permutations to a series  $Q(a, u)$  of quarter-plane walks:

$$2P(t) - 1 = Q\left(\frac{1}{P} - 1, \frac{tP^2}{(1 - 2P)^2}\right).$$

They also considered *primitive* permutations, which are permutations which cannot be written as a sum  $\pi \oplus \tau$  where  $\pi$  and  $\tau$  are non-empty permutations. Equivalently to the equation above, Albert and Bousquet-Mélou showed that the generating function  $p(t)$  of primitive 2sip-sortable permutations satisfies

$$\frac{1 + p(t)}{1 - p(t)} = Q\left(-p, \frac{t}{(1 + p)^2}\right),$$

where  $p(t)$  and  $P(t)$  are related by the equation

$$P(t) = \frac{1}{1 - p(t)}.$$

It will be convenient to compute coefficients of  $p(t)$  and use those to compute coefficients of  $P(t)$  rather than computing coefficients of  $P(t)$  directly. Note that  $p(t)$  is called  $S^\bullet$  in [3].

The first step in the algorithm is to compute coefficients of  $Q(a, u)$ . For non-negative integers  $n, k, x, y$ , let  $s(n, k, x, y)$  denote the number of  $n$  step quarter-plane walks which start at  $(0, 0)$  and end at  $(x, y)$ , with  $k$  ES or NW corners. Then

$$Q(a, u) = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} s(2n, k, 0, 0) a^k u^n.$$

First we calculate  $s(n, k, x, y)$  for each  $n \leq 2N$  and  $x, y \leq \min\{n, 2N - n\}$  via the recurrence relation

$$\begin{aligned} s(n, k, x, y) = & s(n-1, k, x-1, y) + s(n-1, k, x, y-1) \\ & + s(n-1, k, x+1, y) + s(n-1, k, x, y+1) \\ & + s(n-2, k-1, x+1, y-1) - s(n-2, k, x+1, y-1) \\ & + s(n-2, k-1, x-1, y+1) - s(n-2, k, x-1, y+1). \end{aligned}$$

As  $k \leq n/2$  for all non-zero values of  $s(n, k, x, y)$ , the number of terms  $s(n, k, x, y)$  which need to be calculated is a quartic function of  $N$ , so this

algorithm takes quartic time, assuming that addition takes constant time. However, to calculate all of the terms  $s(n, k, x, y)$  for a specific value of  $n$ , only the terms  $s(n-1, k, x, y)$  and  $s(n-2, k, x, y)$  are required, so it is only necessary to keep terms in memory for at most three consecutive values of  $n$  at the same time. Hence, the memory requirements of this algorithm are cubic in  $N$ .

Now that we have calculated coefficients of  $Q(a, u)$ , we can expand

$$F(p, t) = Q\left(-p, \frac{t}{(1+p)^2}\right) - \frac{1+p}{1-p}$$

as a series in  $p$  and  $t$ . Next, calculate the first  $N$  terms of the series  $p(t)$  which sends  $F(p(t), t)$  to 0. Finally we calculated the first  $N$  terms of  $P(t)$ , using the equation

$$P(t) = \frac{1}{1-p(t)}.$$

Using the 2sip series, along with Theorem 2.19, one can quickly generate  $N$  coefficients of the deque series  $D(t)$ .

Calculating the series  $D(t)$  and  $P(t)$  from the terms in  $Q(a, u)$  is very fast, so both the time and space requirements of this algorithm are determined by the efficiency of calculating the coefficients of  $Q$ . We saw that this was quartic in time and cubic in memory, however this did not take into account that the number of digits required to store each number grows on average linearly in  $N$ , as the numbers themselves grow exponentially. If this algorithm was implemented as described, both the time and memory requirements would be multiplied by a linear factor to account for this fact. To avoid this increased memory consumption, one can run this entire algorithm modulo primes  $p \approx 2^{63}$ , then compute the coefficients of the series using the Chinese Remainder Theorem, as long as the product of all primes used is larger than coefficient  $[t^N]P(t)$ . The number of primes required is therefore linear in  $N$ , so the time requirement of this algorithm is in fact quintic in  $N$ . As the algorithm can be run separately for each prime, the memory requirement remains cubic.

The series start:

$$P(t) = 1 + t + 2t^2 + 6t^3 + 23t^4 + 103t^5 + 513t^6 + 2760t^7 + 15741t^8 + \dots$$

$$D(t) = 1 + t + 2t^2 + 6t^3 + 24t^4 + 116t^5 + 634t^6 + 3762t^7 + 23638t^8 + \dots$$

In Section 2.5 we present an empirical analysis of these two counting sequences using differential approximants. In Section 2.6 we use these exact terms to deduce improved lower bounds on the growth rate of both deque-sortable permutations and 2sip-sortable permutations. This also leads to an improved lower bound on the growth rate of 2sis-sortable permutations.



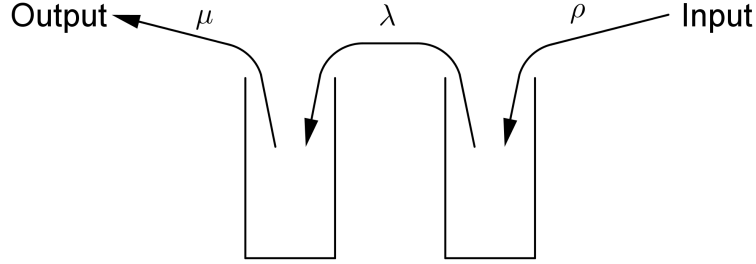


Figure 2.6: Two stacks in series

## Permutations sortable by two stacks in series

In this section we describe the algorithm we have used to evaluate the exact number of 2sis-sortable permutations of length  $n$  for  $n < 20$ . This is a computationally difficult problem. The only existing series we can find in the literature for  $n \leq 10$  are in the PhD thesis of Pierrot [167], and the last two values are incorrect. We also note that Andrew Conway independently calculated the terms for  $n \leq 10$ , and these match our computations.

In Section 2.5, we present empirical analysis of the series using differential approximants. This analysis suggests that the coefficients of the generating function  $S(t)$  of 2sis-sortable permutations behave as

$$s_n \sim a \cdot \mu^n n^g,$$

where  $\mu \approx 12.4$ ,  $g \approx -2.5$  and  $a \approx 0.008$ .

### Basic algorithm

We start with a simple, but inefficient algorithm to calculate the coefficients of the OGF  $S(t)$ , on which our more efficient algorithm is based. Consider the three moves  $\rho$ , which pushes the next element from the input onto the first stack,  $\lambda$ , which pushes the top element of the first stack onto the second stack, and  $\mu$ , which outputs the top element of the second stack as shown in Figure 2.6. We have already defined sortable permutations. We call a permutation of length  $n$  *achievable* if it is possible to output that permutation, starting with the numbers  $1, 2, \dots, n$  in order. Rather than enumerating sortable permutations directly, we will instead enumerate achievable permutations, since the two classes share the same OGF. We call a word  $w$  over the alphabet  $\{\rho, \lambda, \mu\}$  an *operation sequence* if  $w$  corresponds to a permutation. That is,  $w$  is called an operation sequence if  $w$  contains an equal number of occurrences

of each of the three letters, and after any point in  $w$ , the letter  $\rho$  has appeared at least as many times as  $\lambda$ , which has appeared at least as many times as  $\mu$ . Call two operation sequences equivalent if they produce the same permutation. Note that this also means that they sort the same permutation. The basic algorithm, which we will call algorithm 1, works as follows:

- Define the function *addreachableperms* which takes in the state  $S$  of the sorting machine, and a set of permutations and adds every permutation which can be achieved from that state to the set, by recursively calling the same function on each of the three or fewer states which can be reached from  $S$  by one of the moves  $\rho$ ,  $\mu$ , or  $\gamma$ .
- create an empty set  $P$  of permutations.
- Call the function *addreachableperms* on the initial state of the stack and the set  $P$
- Then the  $n$ th coefficient of the OGF is equal to the size of  $P$ , since the permutations in  $P$  are exactly the achievable permutations of size  $n$ .

This algorithm is very slow because it has to consider all operation sequences of size  $3n$  separately, and the number of operation sequences of length  $3n$  grows like  $27^n$ .

## Forbidden words and regular languages

The first improvement which we make is to reduce the number of operation sequences which the algorithm has to consider by removing many operation sequences which create the same permutation. Call two operation sequences equivalent if they create the same permutation. We define the ordering  $\rho < \lambda < \mu$ , and we call an operation sequence optimal if it is lexicographically larger than any other equivalent operation sequence. Rather than parse all operation sequences of size  $3n$ , we now only insist that we parse all optimal operation sequences, since these will still create all achievable permutations. Call a word  $v$  over the alphabet  $\{\rho, \lambda, \mu\}$  forbidden if there is another word  $v' > v$ , which has the same effect on the sorting machine as  $v$ . Note that if an operation sequence  $w$  has a forbidden subword  $v$ , then we can change  $v$  to  $v'$  in  $w$  to create an equivalent operation sequence  $w'$ . Moreover,  $w' > w$ , so  $w$  is not optimal. Hence, any optimal operation sequence contains no forbidden words. Note that  $\rho\mu$  is a forbidden word, since it has the same effect on the sorting machine as  $\mu\rho$ . Also,  $\rho\lambda\mu\lambda$  is a forbidden word since it has the same effect on the sorting machine as  $\lambda\rho\lambda\mu$ . For letters  $x$  and  $y$ , we call a word

$v$  over the alphabet  $\{x, y\}$  an  $x, y$ -Catalan word if the following conditions hold:

- $v$  contains an equal number of  $x$ 's and  $y$ 's
- for any leading subword  $u$  of  $v$ , the word  $u$  contains at least as many  $x$ 's as  $y$ 's.

In other words, if we replace each  $x$  in  $v$  with an up step and each  $y$  in  $v$  with a down step, we get a Dyck path. Note that if  $u$  is a  $\rho, \lambda$ -Catalan word, and  $v$  is a  $\lambda, \mu$ -Catalan word, then the effect of  $u$  on the sorting machine is to move and permute items from the input to the second stack. The effect of  $v$  is to move and permute items from the first stack to the output. Hence, these two operations commute, so  $uv$  and  $vu$  are equivalent. Since  $u$  begins with  $\rho$  and  $v$  begins with  $\lambda$ , we have  $uv < vu$ , so  $uv$  is a forbidden word.

We now construct the deterministic infinite state automaton  $\Gamma$  shown in Figure 2.7, which accepts all words which are not forbidden, as well as some words which are forbidden. For an operation sequence  $w$  of size at most  $3n$ , the word  $w$  is accepted by  $\Gamma$  if and only if  $w$  does not contain any of the words  $\rho\mu$ ,  $\rho\lambda\mu\lambda$  or any word of the form  $uv$ , where  $u$  is a  $\rho, \lambda$ -Catalan word, and  $v$  is a  $\lambda, \mu$ -Catalan word. Since all of these words are forbidden, any operation sequence  $w$  which is not forbidden is accepted by  $\Gamma$ . For any integer  $m$ , at least  $m$  occurrences of the letter  $\rho$  are required to reach any of the states  $a_m$ ,  $b_m$  or  $c_m$ , and at least  $m + 2$  occurrences of the letter  $\lambda$  are required to reach the state  $d_m$ . hence, for operation sequences of size  $n$ , we only need to construct the finite state automaton  $\Gamma_n$ , consisting of the  $4n$  states  $s_0, s_1, a_1, \dots, a_n, b_1, \dots, b_n, c_1, \dots, c_n, d_1, \dots, d_{n-2}$ .

Next, we create two functions to construct a more restrictive DFA  $\Gamma'_n$ , which still accepts all optimal operation sequences. The first function inputs a list of words over the alphabet  $\{\rho, \lambda, \mu\}$  and outputs a DFA which accepts the regular language of all words which do not contain any elements of the input list as subwords. This DFA contains one state for each possible suffix of length at most  $k - 1$ , where  $k$  is the maximum length amongst words in the input list. The second function inputs two DFAs and outputs the DFA which accepts the intersection of the languages accepted by each of the two input DFAs. Now, using a simple brute force algorithm, we make a list  $l$  of all forbidden words of length at most 14 which contain no other forbidden words, and which are accepted by  $\Gamma$ . Using the first function, we construct the DFA  $\Gamma_l$  which accepts the language of all words which do not contain a subword in  $l$ . Then every word which is not accepted by  $\Gamma_l$  contains some forbidden word  $u$  in  $l$ , and is hence forbidden. We then use the second function to construct the DFA  $\Gamma'_n$ , which accepts the intersection of the languages accepted by  $\Gamma_l$

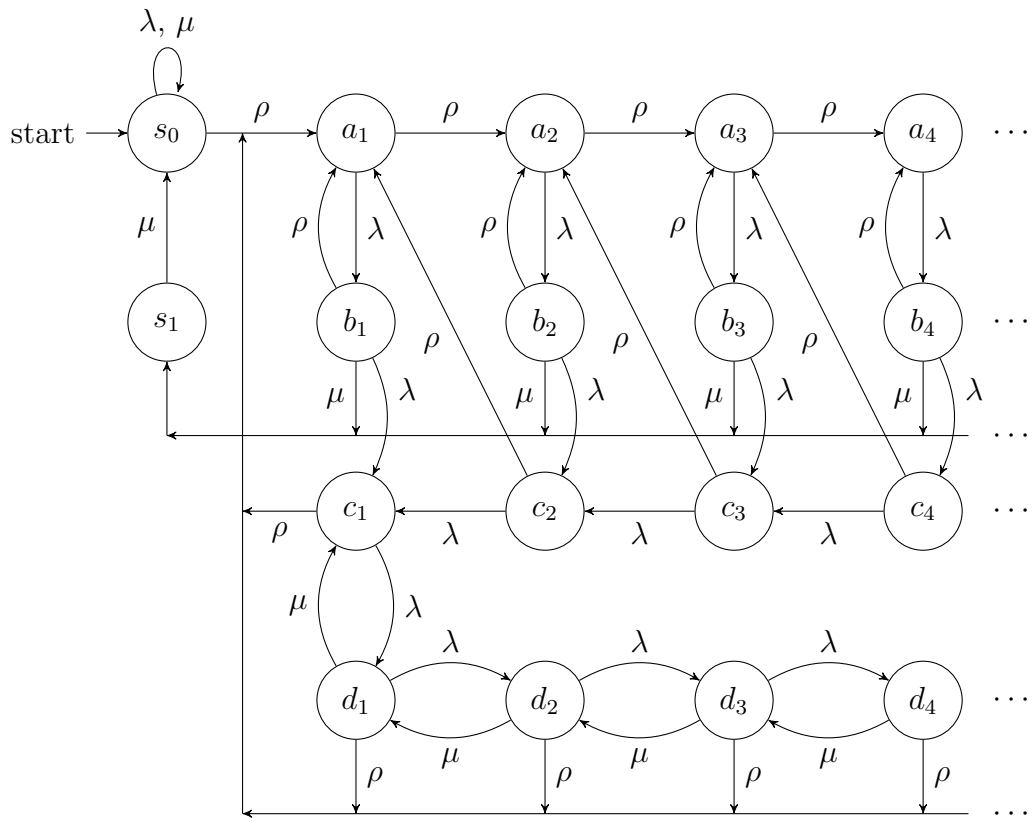


Figure 2.7: The infinite state automaton  $\Gamma$

and  $\Gamma_n$ . In total there are 207 words in  $l$ , made up of 2, 4, 8, 13, 22, 81 and 77 words of lengths 7, 9, 10, 11, 12, 13 and 14, respectively. The two of length 7 are  $\rho\lambda\rho\lambda\mu\mu\lambda$  and  $\rho\lambda\lambda\rho\lambda\mu\mu$ , which are forbidden since they have the same effect on the sorting machine as  $\lambda\rho\lambda\rho\lambda\mu\mu$  and  $\lambda\rho\rho\lambda\mu\mu\lambda$ , respectively.

Now, our new algorithm works the same way as in Section 2.4.1, except that the function *addreachableperms* also takes in the current state  $A$  of  $\Gamma'_n$ , and only recursively calls itself using one of the letters which is accepted from state  $A$ . Now, rather than considering all operation sequences of size  $3n$ , the new algorithm only considers those operation sequences which are accepted by  $\Gamma'_n$ .

The improvements to the algorithm so far significantly decrease the exponential factor in the time requirement, from 27 to about 13. However the algorithm still stores every achievable permutation of length  $n$  in memory at the same time. In the Section 2.5 we see that the number of such permutations is approximately  $12.4^n$ , so any improvements of the form which we have presented so far will not reduce the time or memory requirements below this factor.

## Increment-avoiding permutations

Our next improvement to the algorithm decreases the exponential factor by 1, and we do not improve on the factor for time any more than this. Let  $p = a_1 \dots a_n$  be a permutation and let  $I \subset \{1, 2, \dots, n\}$ . We define the subpermutation  $p|_I$  to be the pattern of the elements from  $I$  in  $p$ , and define  $\hat{p}|_I = p|_{\{1, \dots, n\} \setminus I}$ . For example,  $\widehat{24315}|_{\{2,3\}} = 24315|_{\{1,4,5\}} = 213$ . Note that any subpermutation of an achievable permutation is also achievable. Let  $p$  be a permutation such that  $a_{j+1} = a_j + 1$  for some  $j$  (sometimes called an increasing bond). Since  $p$  is achievable,  $\hat{p}|_{\{a_j\}}$  is achievable. On the other hand, if  $\hat{p}|_{\{a_j\}}$  is achievable, then there is some operation sequence  $w$  which creates it. Now replace each letter which moves  $a_j$  in  $w$  with two copies of that letter, to form a new word  $w'$ . Then the two copies of each letter will move  $a_j$  and  $a_j + 1$ , and  $a_j$  will enter the first stack immediately before  $a_j + 1$ , then  $a_j + 1$  will enter the second stack immediately before  $a_j$  and finally,  $a_j$  is output immediately before  $a_j + 1$ . Since the order of everything else stays the same, the word  $w'$  creates the permutation  $p$ . Therefore,  $p$  is achievable if and only if  $\hat{p}|_{\{a_j\}}$  is achievable.

Now, instead of considering all achievable permutations with the algorithm, we only consider permutations  $a_1 \dots a_n$  for which there is no  $j$  satisfying  $a_{j+1} = a_j + 1$ . Call these increment-avoiding permutations (sometimes called plus-irreducible permutations). Let  $r_n$  be the number of these permutations, and define the generating function  $R(t) = r_0 + tr_1 + \dots$ . Then we can

uniquely create any 2sis-achievable permutation by choosing a permutation  $\pi$  counted by  $R$  and replacing each number in  $\pi$  with any positive number of consecutive integers. Hence  $S(t) = R(t/(1-t))$ . By taking the coefficient for  $t^n$  on both sides of this equation, we deduce that

$$s_n = \sum_{i=1}^n \binom{n-1}{i-1} r_i.$$

The only change we make to the algorithm presented previously is to instead calculate the number of increment-avoiding permutations is to forbid an item from being output if it is exactly one greater than the previous item output.

## Memory consumption and parallelisation

Using the algorithm described so far, it is still necessary to list every achievable, increment-avoiding permutation of size  $n$  at the same time. To avoid this restriction, we choose some positive integer  $m < n$ , and write a function *numpermswithprefix*, which inputs  $n$  and a sequence  $s$  of  $m$  distinct elements of  $\{1, \dots, n\}$  and outputs the number of achievable, increment-avoiding permutations of size  $n$  which begin with the sequence  $s$ . This algorithm works in the same way as before except that the first  $m$  elements output must be the correct elements of  $s$ . We then run this function on all such sequences  $s$  and add up the results. For small values of  $m$  this algorithm only takes a little longer than the original algorithm because most of the time is spent while the operation sequence is long and the output is nearly complete. Since we call the function on different sequences  $s$  separately, it is only necessary to store all of the (achievable, increment-avoiding) permutations which begin with some sequence  $s$  at any one time. Note also that we only have to remember the last  $n - m$  elements of each permutation. As a result, the limiting factor for this algorithm is now the time requirement.

We now parallelise the algorithm, by running the *numpermswithprefix* on different sequences  $s$  at the same time on different cores.

## Results

We ran this algorithm for  $n < 20$  using  $m = 6$ . The program ran for 43 days on 64 cores. The coefficients of the OGF for  $n < 20$  are given as a list below.

[1, 1, 2, 6, 24, 120, 720, 5018, 39374, 337816, 3092691, 29659731, 294107811, 2988678546, 30935695794, 324832481490, 3450158410649, 36993206191004, 399827092167771, 4351269802153188].

## Empirical analysis using differential approximants

In this section we make a numerical study of the three generating functions  $P(t)$ ,  $D(t)$  and  $S(t)$ , for 2sip-sortable permutations, deque-sortable permutations and 2sis-sortable permutations. We use the two most common methods of series analysis, the ratio method and the method of differential approximants. Full details of these methods can be found in, for example, [122]. Both methods aim to estimate the radius of convergence (r.c.)  $z_c$  and associated exponent  $\theta$  of functions whose asymptotic behaviour is given by

$$F(z) \sim A \left(1 - \frac{z}{z_c}\right)^\theta, \text{ as } z \rightarrow z_c^-. \quad (2.11)$$

It follows that

$$f_n = [z^n]F(z) \sim \frac{An^{-\theta-1}}{\Gamma(-\theta)z_c^n}. \quad (2.12)$$

### Ratio method.

The ratio method, as the name implies, relies on extrapolating the ratio of successive coefficients,  $r_n$ . One has

$$r_n = \frac{f_n}{f_{n-1}} = \frac{1}{z_c} \left(1 - \frac{\theta+1}{n} + o(1/n)\right). \quad (2.13)$$

Clearly, extrapolating the ratios  $r_n$  against  $1/n$  should, for sufficiently large  $n$ , give a linear plot which extrapolates to  $1/z_c$  at  $1/n = 0$ . The gradient is  $-(\theta+1)/z_c$ . So from the intercept one can estimate the radius of convergence, and from the gradient and the estimate of the radius of convergence, one can estimate the exponent  $\theta$ . In favourable cases, where the singularity is precisely as given in (2.11), the term  $o(1/n)$  can be replaced by  $O(1/n^2)$ , and the ratio plot will usually be linear from quite low values of  $n$ .

If the r.c. is known, or very accurately estimated from some other method, it follows from (2.13) that a more precise estimate of the exponent can be made from estimators

$$\theta_n = n(1 - z_c \cdot r_n) - 1 + o(1), \quad (2.14)$$

where in favourable cases the term  $o(1)$  can be replaced by  $O(1/n)$ .

Even if the r.c. is not known, one can obtain an estimate of the exponent independent of the r.c. by extrapolating ratios of ratios, so that

$$t_n = \frac{r_n}{r_{n-1}} = \left(1 + \frac{1+\theta}{n^2} + o\left(\frac{1}{n^2}\right)\right). \quad (2.15)$$

So we can define  $\theta_n$ , an estimator of the exponent  $\theta$ , as

$$\theta_n = (t_n - 1)n^2 - 1 = \theta + o(1). \quad (2.16)$$

where again, in favourable cases, the term  $o(1)$  can be replaced by  $O(1/n)$ .

## Method of differential approximants

The method of differential approximants [122] fits the known coefficients of a power series to a number (typically 10 or 12) of holonomic differential equations, and uses the critical parameters (the radius of convergence and exponent at that point) of those differential equations as estimators of the corresponding quantities for the underlying series expansion.

More precisely, one uses the known series coefficients to find polynomials  $Q_k(z)$  and  $P(z)$  such that the power series solution  $\tilde{F}(z)$  of the holonomic differential equation

$$\sum_{k=0}^M Q_k(z) \left(z \frac{d}{dz}\right)^k \tilde{F}(z) = P(z) \quad (2.17)$$

agrees with the known coefficients of the function  $F(z)$  being approximated. The order  $M$  of the ode we refer to as the *order* of the approximant.

Constructing such *differential approximants* (DAs) is straightforward computationally, and only involves solving a linear system of equations. Several such DAs are constructed by varying the degrees of the polynomials  $Q_k(z)$  and  $P(z)$ , while still using most, or all of the known series coefficients. The singularities are given by the zeros  $z_i$ ,  $i = 1, \dots, N_M$  of  $Q_M(z)$ , where  $N_M$  is the degree of  $Q_M(z)$ . We take as the dominant singularity that which is both closest to the origin and common to all (or almost all) the DAs. Critical exponents  $\theta_i$  follow from the indicial equation of the DA. Our predictions of these critical exponents assumes a regular singular point. For the simplest (and most frequent) situation where there is a single root of  $Q_M(z)$  at  $z_i$ ,

$$\theta_i = 1 - M + \frac{Q_{M-1}(z_i)}{z_i Q'_M(z_i)}.$$

Slightly more complicated expressions are known for the cases of double, triple etc. roots. Further details of both methods can be found in [122].



## Analysis of series for dequeues and 2sip

We subjected the first 500 terms of the series  $P(t)$  for 2sip-sortable permutations and  $D(t)$  for deque-sortable permutations to differential approximant analysis.

For 2sips, 6th order DAs show the dominant singularity to be at  $t_c = 0.1207524975763(2)$  with an exponent  $1.47309(3)$ , with another exponent with the value  $1.94652(2)$  at  $t_c = 0.1207524975763(1)$ . A third exponent at the same place with the value  $4.72(4)$  is also suggested. The quoted errors reflect only the scatter in individual approximant estimates, and because of the presence of confluent singularities should be multiplied by a factor of 10 at least to be on the safe side.

With 8th order DAs, we find the dominant singularity to be at  $t_c = 0.1207524975763(2)$  with an exponent  $1.47309(4)$ , and with another exponent  $1.94652(2)$  at  $t_c = 0.1207524975764(8)$ . A third exponent at the same place with the value  $4.72(4)$  is also suggested.

For the deque series, the unbiased analysis showed the dominant singularity to be at  $z_c = 0.12075249773(4)$ , with exponents values of  $0.507(4)$ ,  $0.970(2)$ , and  $1.414(1)$ . The uncertainties quoted just reflect the variability in the estimates across many DAs, and the lack of overlap between  $z_c$  estimates for dequeues and 2sips beyond the 10th significant digit suggests that they are too optimistic. Nevertheless, 10-digit agreement gives considerable credence to the conjecture that they are indeed equal, and we assume this for the subsequent analysis. The dominant exponent is very close to  $1/2$  exactly, which provides numerical support for the conjectured square-root singularity obtained in the previous section.

Another way to study the exponents for these two problems is by the ratio method. We write  $[t^n]P(t) \sim \text{const} \cdot \mu^n \cdot n^{g_p}$  and  $[t^n]D(t) \sim \text{const} \cdot \mu^n \cdot n^{g_d}$ . Then from the Hadamard (coefficient-by-coefficient) quotient,

$$q_n = \frac{[t^n]P(t)}{[t^n]D(t)} \sim \text{const} \cdot n^\theta,$$

where  $\theta = g_p - g_d$ . Then the ratios of successive coefficients  $r_n$  behave as

$$r_n = \frac{q_n}{q_{n-1}} \sim 1 + \frac{\theta}{n} + o(1/n).$$

We can estimate the exponent  $\theta$  by extrapolating a sequence of estimators  $\{\theta_n\}$ , defined by  $n \cdot (r_n - 1) \sim \theta_n + o(1)$ . The result is shown in Figure 2.8, in which  $\theta_n$  is shown extrapolated against  $1/n$ . While it is difficult to extrapolate this curve, a limit in the range  $[-0.975, -0.972]$  looks plausible.

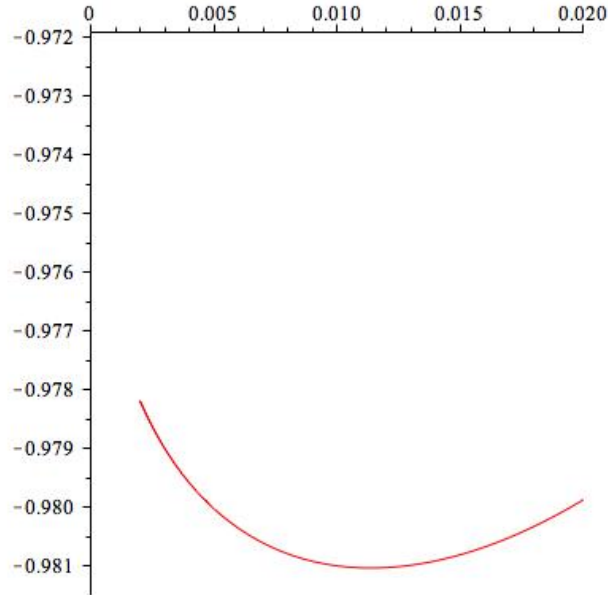


Figure 2.8: Exponent estimates  $\theta_n$  plotted against  $1/n$ . Note limit appears to be around -0.974.

We also analysed the deque series by the ratio method. From eqn (2.16), estimators of the deque exponent  $g_d$  can be found, and these are shown extrapolated against  $1/n$  in Fig 2.9. The plot is quite linear and is clearly going to a value close to  $-1.5$ .

We can refine this by calculating the exponent  $g$  more precisely from the local gradient of the previous curve. This is  $h_n = (g_n - g_{n-1})n(1 - n)$ , where  $g_n$  is the  $n^{\text{th}}$  estimator of  $g_d$ . Then a straight line with this gradient will meet the ordinate at  $g_n - \frac{h_n}{n}$ , which should be a refined estimator of the exponent  $g$ . This plot is shown in Fig 2.10. That the curve appears to be going slightly below  $-1.5$  is, we believe, of no consequence. We believe that if we had several thousand terms we would see this curve pass through a minimum, and increase to  $-1.5$  exactly.

Accepting the conjectured square-root singularity of the deque generating function, numerical results thus far suggest that

$$D(t) = D(t_c) + b_1 \sqrt{1 - t/t_c} + b_2(1 - t/t_c)^{\alpha-1/2} + \dots \quad (2.18)$$

where  $1 < \alpha \approx 1.47 < 3/2$ , as this gives the value 0.97 observed as the sub-dominant exponent for deques. Similarly, it appears that

$$P(t) = P(t_c) + c_1(1 - t/t_c) + c_2(1 - t/t_c)^\alpha + o((1 - t/t_c)^\alpha), \quad (2.19)$$

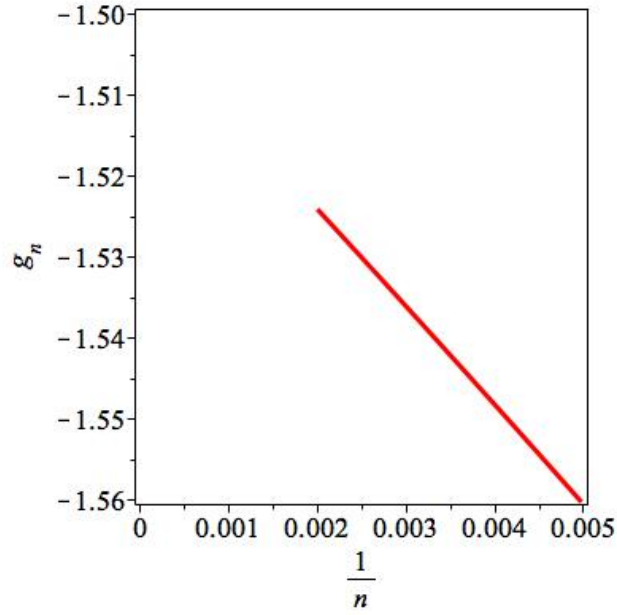


Figure 2.9: Exponent estimates  $g_n$  plotted against  $1/n$  for deques. The limit appears to be  $g \approx -1.5$ .

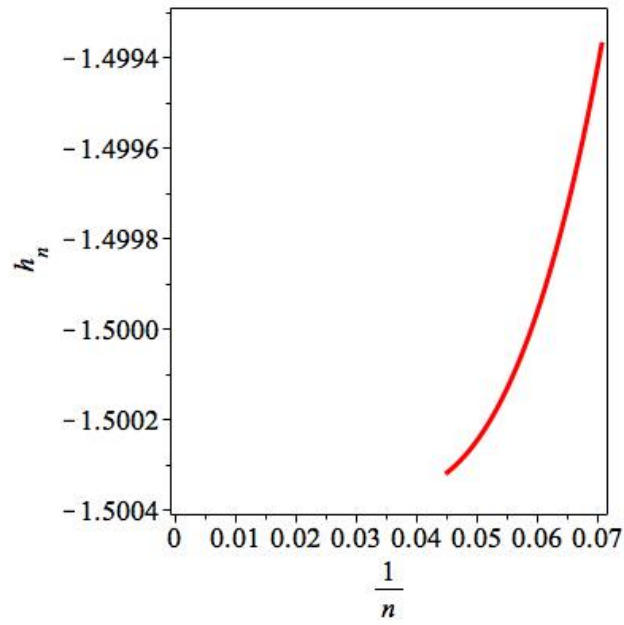


Figure 2.10: Extrapolated exponent estimates  $h_n$  plotted against  $1/n$  for deques. The limit appears to be around  $g \approx -1.500$ .

as predicted in Theorem 2.26. Recall our prediction from Theorems 2.25 and 2.26 that

$$D(t_c) = 2 + t_c + 2c_0t_c - 2c_0t_c^2, \quad (2.20)$$

$$b_1 = -t_c \sqrt{(2c_0)^{3/2}t_c - 2c_1\sqrt{t_c}} \quad (2.21)$$

and

$$b_2 = \frac{-t_c^{3/2}c_2}{\sqrt{(2c_0)^{3/2}t_c - 2c_1\sqrt{t_c}}}, \quad (2.22)$$

where  $c_0 = P(t_c)$ .

We also estimated the value of  $c_1$  numerically from Padé approximants to the series for  $P(t) - P(t_c)$ , evaluated at  $t = t_c$ . Numerical estimates of  $c_1$  give  $b_1 \approx -0.0540$  from (2.21), which is precisely equal to direct estimates of  $b_1$  obtained from our numerical analysis of the deque series. There we formed the series for

$$\frac{D(t) - D(t_c)}{\sqrt{1 - t/t_c}}$$

and evaluated the approximants at  $t = t_c$ . Furthermore, using this amplitude value  $b_1$ , and subtracting  $b_0 + b_1\sqrt{1 - t/t_c}$  from the deque series  $D(t)$  gives a remainder series that behaves as  $\text{const} \cdot (1 - t/t_c)^\theta$ , where  $\theta \approx 0.973$ . So both ratio and differential approximant analyses clearly identify this confluent exponent.

Similarly, we analysed the series for two stacks in parallel by the ratio method. The exponent estimators  $g_n$ , given by eqn (2.14), are plotted against  $1/n$  in Figure 2.11. The plot is visually quite straight and is clearly going to a limit of about -2.475, as shown in Fig 2.11.

As with the deque series, we can refine this exponent estimate by calculating the exponent  $g$  more precisely from the local gradient of the previous curve. The refined estimate of the exponent  $g$  is shown in Fig 2.12, and from that plot a limit around -2.474 seems quite plausible.

So the ratio plots support the conclusion that the exponent for the deque generating function is very close to 0.5, corresponding to a square-root singularity, as conditionally proved in Theorem 2.25. The first confluent exponent for deque has the approximate value 0.973, and these two exponents add to give the observed value 1.473 for the exponent of the ogf for 2sips, as predicted by Theorem 2.25. A direct ratio analysis of the ogf for 2sips also gives an exponent value around 1.472 – 1.474.

Finally, we calculate the amplitudes, or pre-multiplying constants  $\kappa_d$  and  $\kappa_p$ , where

$$[t^n]D(t) = d_n \sim \kappa_d \cdot t_c^{-n} \cdot n^{-3/2}, \text{ and } [t^n]P(t) = p_n \sim \kappa_p \cdot t_c^{-n} \cdot n^{-2.473}.$$

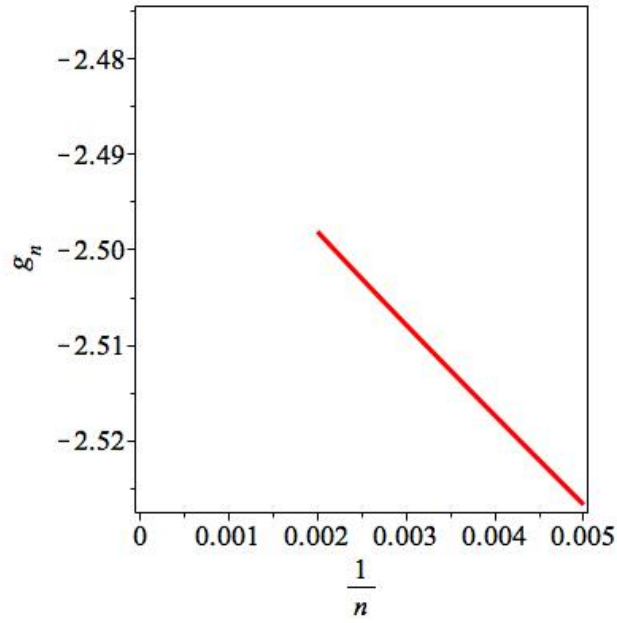


Figure 2.11: Exponent estimates  $g_n$  plotted against  $1/n$  for two stacks in parallel. The limit appears to be about  $g \approx -2.475$ .

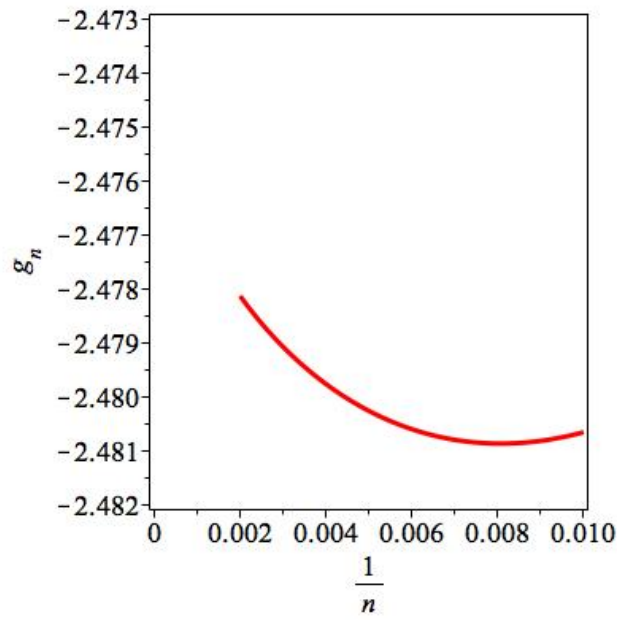


Figure 2.12: Extrapolated exponent estimates  $g_n$  plotted against  $1/n$  for two stacks in parallel. The limit appears to be around  $g \approx -2.474$ .

We do this by forming the sequences  $d_n \cdot t_c^n \cdot n^{3/2} \sim \kappa_d + o(1)$  and  $p_n \cdot t_c^n \cdot n^{2.473} \sim \kappa_p + o(1)$ , and extrapolating these against  $1/n$ . In this way we estimate  $\kappa_d = 0.01524 \pm 0.0005$ , and  $\kappa_p = 0.08025 \pm 0.0010$ .

Moving now from the arena of careful numerical work to that of wild speculation, in some unpublished work we have studied the behaviour of 421-3 pattern-avoiding permutations. We identified the critical exponent numerically, as

$$\frac{2}{3} \left( 1 + \frac{2\pi}{3\sqrt{3}} \right) = 1.472799717437 \dots$$

We were struck by the fact that this is tantalisingly close to the observed exponent of two stacks in parallel, though there is no obvious reason that the two problems should be connected, except that they both involve pattern-avoiding permutations.

## Series extension for 2sis.

Recall that we computed the coefficients  $s_n$  of the generating function  $S(t)$  for  $n < 20$ . We have obtained approximate values of the next nineteen coefficients, effectively doubling the length of the series, which are sufficiently accurate to be used in the ratio analysis we describe below. Our method for obtaining these approximate values uses *differential approximants* [122], which, as discussed in Section 2.5.2, are linear, inhomogeneous ODE's of 2nd, 3rd or 4th order, constructed to yield all the exactly known coefficients in the series expansion under consideration. By varying the degrees of the polynomials multiplying each derivative, as well as the degree of the inhomogeneous polynomial, we can construct a family of such approximants. Because every differential approximant (DA) implicitly predicts all subsequent coefficients, we can calculate, approximately, all subsequent coefficients. Of course the accuracy of these predicted coefficients decreases as the order of the predicted coefficients increases, but, as we show by example below, we can get useful estimates of the next nineteen or so coefficients.

For every DA using all known coefficients, we generated the subsequent nineteen coefficients. We take the mean of the predicted coefficients, with the outlying 10% or 15% of estimates rejected, as our estimate. We quote one standard deviation as the error. That is to say, using the coefficients  $a_n$  for  $n \in [0, 19]$ , we then predict the coefficients  $a_n$  for  $n \in [20, 38]$ . Our estimate of each such coefficient is given by the mean of the values predicted by the differential approximants. We reject obvious outliers, by discarding the top and bottom 10% of estimates. Not surprisingly, we find the smallest error is predicted for  $a_{20}$ , with the error slowly increasing as we generate further coefficients.

These predicted coefficients are well-suited to ratio type analyses, as discrepancies in say the seventh or eighth significant digit will not affect the ratio analysis in the slightest. This is particularly useful in those situations where we suspect there might be a turning point in the behaviour of ratios or their extrapolants with our exact coefficients, as these approximate coefficients are more than accurate enough to reveal such behaviour, if it is present.

As an indication of the validity of this method, we give two applications. In the first, we take the series  $P(t)$  for two stacks in parallel, for which we actually have more than 1000 coefficients, but assume we only have the first 20 coefficients, just as in the present case for the generating function of two stacks in series. In Table 2.1 we show a selection of the estimated coefficients  $p(20)$  to  $p(38)$ . It can be seen that we predict the next coefficient with an accuracy of 13 digits, decreasing to 7 digit accuracy for the last predicted coefficient. In every case the actual error is seen to be less than one standard deviation, indeed, it is typically 1/3 of a standard deviation.

$N$	$p_N$ estimate	1 std. devn.	actual error
20	$1.36000505625858 \times 10^{14}$	81	28
21	$9.90406677134907 \times 10^{14}$	5285	1778
22	$7.258100272044 \times 10^{15}$	187074	64212
23	$5.349517582877 \times 10^{16}$	4807109	1358815
24	$3.9634005851 \times 10^{17}$	$9.9546 \times 10^7$	$3.924 \times 10^7$
25	$2.95046460646 \times 10^{18}$	$1.7832 \times 10^9$	$5.767 \times 10^8$
29	$9.435573118 \times 10^{21}$	$7.2824 \times 10^{13}$	$2.222 \times 10^{13}$
32	$4.15469546597 \times 10^{24}$	$1.1960 \times 10^{17}$	$3.569 \times 10^{15}$
35	$1.873198683303 \times 10^{27}$	$1.5136 \times 10^{20}$	$4.030 \times 10^{19}$
38	$8.613038855 \times 10^{29}$	$7.2824 \times 10^{22}$	$2.222 \times 10^{21}$

Table 2.1: Series coefficients  $p_N$  for two stacks in parallel. Approximate, predicted coefficients  $p_{20}$  to  $p_{38}$  from several 4th order inhomogeneous DAs, the estimated and exact error.

As a second demonstration of this method, assume we only have 19 terms in the generating function for two stacks in series, and we'll predict the next coefficient, which is the last known coefficient. The predictions produced by fourth-order DAs are averaged, deleting the top and bottom 10% of estimates. In this way we estimate  $s_{19} = 4.351269803411739 \times 10^{15}$ . The correct answer is 4351269802153188, which is estimated with an error of 1 part in the 10th

significant digit by the differential approximants. The standard deviation of the estimates is 7979922, which is six times the actual error.

In an identical manner to that described above to estimate the coefficients of  $P(t)$  we have obtained estimates of the next 19 coefficients of the generating function  $S(t)$  for two stacks in series, of course using the exact value of  $s_{19}$ . These are given in Table 2.2 below. We also give the standard deviation of the estimates, and based on the examples already discussed, we expect coefficient errors to be less than this. It can be seen that fewer significant digits are predicted than for the two stacks in parallel series—typically 3 or 4 fewer digits at each order. Nevertheless, the precision (4 significant digits at worst), is sufficient for a simple ratio plot.

$n$	$s_n$ estimate	1 std. devn.
20	$4.764211695346 \times 10^{16}$	$9.207 \times 10^6$
21	$5.24460896431 \times 10^{17}$	$9.83 \times 10^8$
22	$5.8016808762 \times 10^{18}$	$7.962 \times 10^{10}$
23	$6.446525027 \times 10^{19}$	$2.241 \times 10^{12}$
24	$7.192361922 \times 10^{20}$	$7.34 \times 10^{13}$
25	$8.05485154 \times 10^{21}$	$2.05 \times 10^{15}$
26	$9.05248613 \times 10^{22}$	$5.10 \times 10^{16}$
27	$1.02070684 \times 10^{24}$	$1.17 \times 10^{18}$
28	$1.15442858 \times 10^{25}$	$2.47 \times 10^{19}$
29	$1.30944006 \times 10^{26}$	$4.95 \times 10^{20}$
30	$1.4893068 \times 10^{27}$	$9.38 \times 10^{21}$
31	$1.6982322 \times 10^{28}$	$1.17 \times 10^{23}$
32	$1.941173 \times 10^{29}$	$2.98 \times 10^{24}$
33	$2.2239807 \times 10^{30}$	$5.05 \times 10^{25}$
34	$2.5535645 \times 10^{31}$	$8.36 \times 10^{26}$
35	$2.938088 \times 10^{32}$	$1.345 \times 10^{28}$
36	$3.387209 \times 10^{33}$	$2.106 \times 10^{29}$
37	$3.91235 \times 10^{34}$	$3.219 \times 10^{30}$
38	$4.52711 \times 10^{35}$	$4.945 \times 10^{31}$

Table 2.2: Series coefficients  $s_n$  for two stacks in series. Approximate, predicted coefficients  $p_{20}$  to  $p_{38}$  from several 4th order inhomogeneous DAs and the estimated error.

If we wish to plot the ratios, we can do better by extrapolating the sequence of ratios produced from the coefficients predicted. That is to say, for each approximating differential approximant one calculates the ratio of



successive coefficients and averages these across all differential approximants using all known coefficients - as usual discarding the outlying 10% or 15% of entries. As shown in [124] this generally gives more accurate ratios than taking ratios of predicted coefficients. In this way we have obtained the next 30 ratios, and these are shown in Table 2.3. We see that we have 10 digit accuracy in the first predicted ratio, decreasing to 4 digit accuracy in the 30th predicted ratio.

## Analysis of extended series for 2sis

We first performed a simple ratio analysis, under the assumption that the coefficients behave as  $s_n \sim \text{const} \cdot \mu^n \cdot n^g$ . Then the ratio of successive coefficients,  $r_n$  behaves as

$$r_n = \frac{s_n}{s_{n-1}} = \mu \left( 1 + \frac{g}{n} + o\left(\frac{1}{n}\right) \right),$$

so plotting the ratios against  $1/n$  should, for sufficiently large  $n$ , give a straight line intercepting the abscissa at  $\mu$ , and with gradient  $g \cdot \mu$ . We show this plot in Figure 2.13. One sees some low  $n$  curvature, and this suggests the presence of a confluent singularity. That is to say, the generating function probably behaves as

$$S(z) \sim A(1 - \mu z)^\gamma + B(1 - \mu z)^{\gamma+\Delta},$$

where  $0 < \Delta < 1$ . Such behaviour implies, at the coefficient level,

$$s_n \sim \frac{A}{\Gamma(-\gamma)} \mu^n n^{-\gamma-1} + \frac{B}{\Gamma(-\gamma-\Delta)} \mu^n n^{-\gamma-\Delta-1},$$

and for the ratios

$$r_n \sim \mu \left( 1 - \frac{\gamma+1}{n} + \frac{\text{const.}}{n^{1+\Delta}} \right).$$

From Figure 2.13 it is seen that  $\mu \approx 12.4$ . Assuming this value, and estimating the gradient from the last plotted point, we find  $g \approx -2.5$ . With  $\mu = 12.3$ , we get  $g = -2.1$ , and with  $\mu = 12.5$ , we get  $g = -2.9$  by this procedure, so it is clear that the estimate of  $g$  is very sensitive to the estimate of  $\mu$ .

Calculating linear intercepts usually gives a more precise estimate of  $\mu$ . One has

$$n \cdot r_n - (n-1) \cdot r_{n-1} = \mu \left( 1 + O\left(\frac{1}{n(1+\Delta)}\right) \right).$$

$n$	$r_n$ estimate	1 std. devn.
20	1.094901468298879e+01	2.11772356e-09
21	1.100834576534045e+01	1.85143285e-08
22	1.106218007359570e+01	8.68930350e-08
23	1.111147816281692e+01	2.96091135e-07
24	1.115695963791318e+01	8.01235631e-07
25	1.119917449576340e+01	1.85222167e-06
26	1.123855154825218e+01	3.82169883e-06
27	1.127543165860235e+01	7.19040042e-06
28	1.131009082195476e+01	1.25541540e-05
29	1.134275634825540e+01	2.05420282e-05
30	1.137361924070142e+01	3.24451987e-05
31	1.140284018941137e+01	4.79461475e-05
32	1.143055939166936e+01	6.84547697e-05
33	1.145689689567404e+01	9.36110721e-05
34	1.148196167309631e+01	1.27067931e-04
35	1.150584576102606e+01	1.65506944e-04
36	1.152863546691212e+01	2.12677243e-04
37	1.155040666870830e+01	2.66928864e-04
38	1.157122649222079e+01	3.29184438e-04
39	1.159116101123194e+01	3.99718853e-04
40	1.161026217312874e+01	4.79900592e-04
41	1.162858140362701e+01	5.64407415e-04
42	1.164616370369958e+01	6.53792233e-04
43	1.166306461316009e+01	7.55633579e-04
44	1.167930152412725e+01	8.55570943e-04
45	1.169494749603971e+01	9.75376673e-04
46	1.171001288824953e+01	1.09733642e-03
47	1.172454051929511e+01	1.22834662e-03
48	1.173852889188502e+01	1.35210096e-03
49	1.175203710982201e+01	1.48442613e-03

*Table 2.3:* Predicted ratios  $r_n = s_n/s_{n-1}$  and their standard deviation for two stacks in series, from 4th order inhomogeneous DAs.

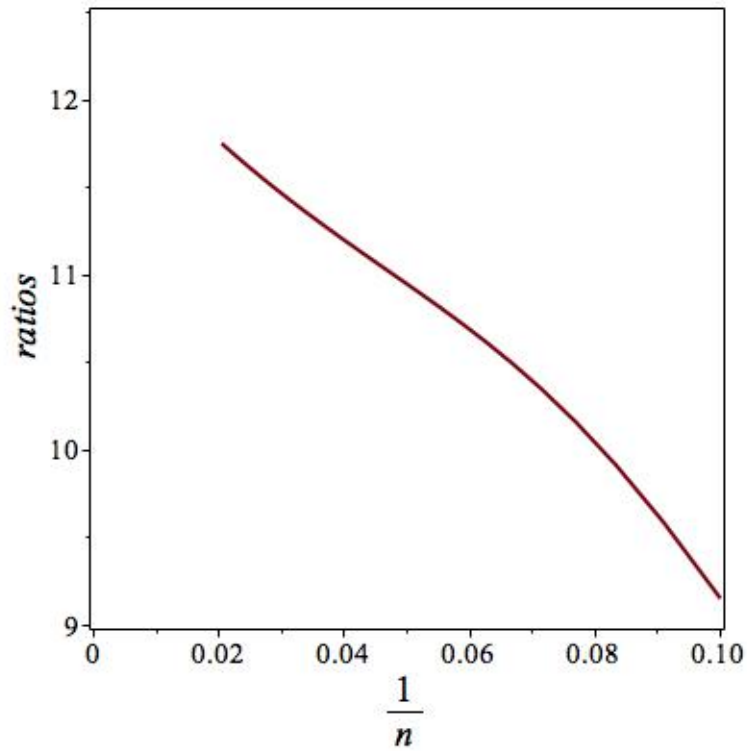


Figure 2.13: Plot of ratios against  $1/n$ .

For a simple power-law singularity, there is no confluent singularity at  $z = z_c = 1/\mu$ , so that  $\Delta = 1$  in the above equation. Then the subdominant term is  $O\left(\frac{1}{n^2}\right)$ , and convergence to  $\mu$  is usually more rapid. However a plot of linear intercepts against  $1/n^2$ , shown in Fig 2.14 has gradient that changes sign for the last few values of  $n$ , and which may change sign again as  $n$  increases, making it difficult to extrapolate, and strongly suggesting the presence of one or more confluent terms. It also implies that we would really need many more series coefficients in order to make more precise estimates of the critical parameters. It also reinforces the usefulness of the sequence extension procedure we have undertaken, as these approximate coefficients are essential to see this change of gradient. Despite these qualifications, a limiting value around  $\mu = 12.4$ , consistent with the value found by a simple ratio plot, seems plausible.

One can also calculate the gradient directly, from

$$\frac{(r_n - r_{n-1}) \cdot n(n-1)}{\mu} = g \left( 1 + o\left(\frac{1}{n}\right) \right).$$

Assuming the values  $\mu = 12.3$ ,  $\mu = 12.4$ , and  $\mu = 12.5$  we have plotted these

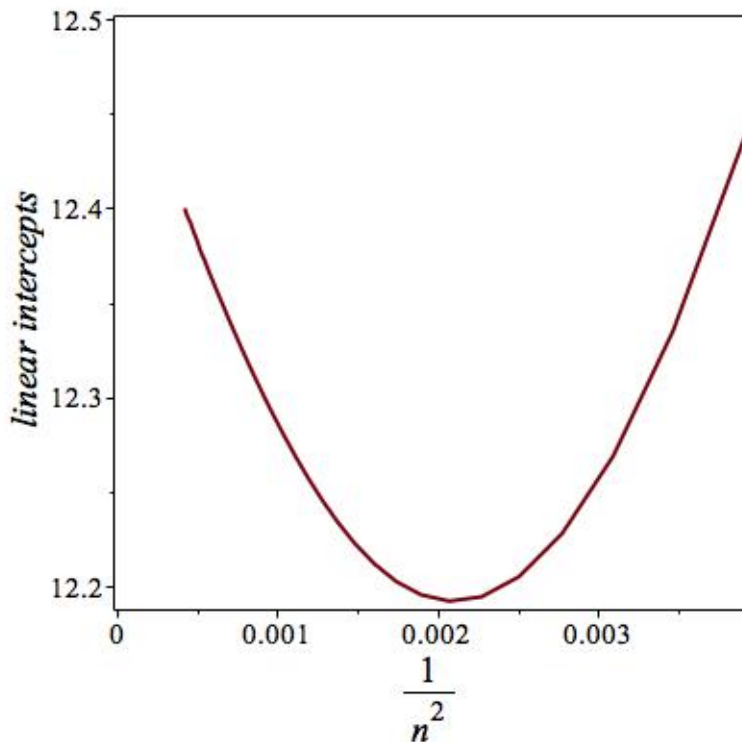


Figure 2.14: Plot of intercepts of successive ratios against  $1/n^2$ .

estimators of  $g$  against  $1/n^2$ , as we don't know the correct sub-dominant exponent to use. Again one sees the necessity of estimating the last few terms, as otherwise the gradient change would not be observed, and a quite inaccurate estimate of  $g$  would be obtained. As it is, we don't know how this plot will behave as  $n$  increases, so cannot give any extrapolation with much confidence. However, if present trends continue, a value of  $g \approx -2.5$  is plausible.

We take as our final estimates  $\mu = 12.45 \pm 0.15$  and  $g_s = -2.5 \pm 0.3$ , where the quoted errors are uncertainty estimates, and not in any sense rigorous error bounds. Alternatively expressed, the OGF for two stacks in series behaves as

$$S(z) \sim \text{const} \cdot (1 - \mu \cdot z)^{-g_s-1} \approx \text{const} \cdot (1 - 12.45z)^{1.5}.$$

Our estimate of  $\mu$  is of course consistent with the rigorous bounds given in [2], which are  $8.156 < \mu < 13.374$ , as well as the improved lower bound 8.183, which we will prove in the next section. It is not inconceivable that the exponent could be the same as for two stacks in parallel, that is, 1.47327, but we have insufficient data to estimate the exponent with anything like

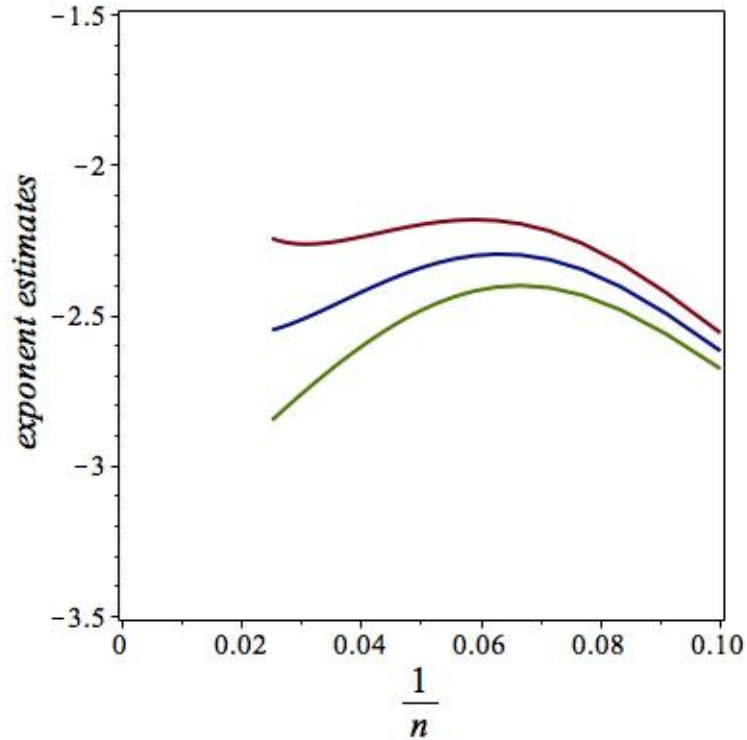


Figure 2.15: Plot of estimators of exponent  $g$  against  $1/n^2$ . The top curve assumes  $\mu = 12.3$ , the middle curve assumes  $\mu = 12.4$  and the bottom curve assumes  $\mu = 12.5$ .

this precision. This would correspond to the two problems being in the same universality class, when viewed from a statistical mechanical perspective.

Assuming the central estimates of both  $\mu$  and the exponent  $g$ , one can estimate the amplitude by simple extrapolation. That is to say, if  $s_n \sim a \cdot \mu^n \cdot n^g$ , then  $a$  can be estimated by extrapolating the sequence  $s_n/(\mu^n \cdot n^g)$  against  $1/n$ . In this way we estimated  $a \approx 0.008$ . Note however that this estimate is very sensitive to the estimates of both  $\mu$  and  $g$ . Writing the singular part of the generating function as  $S(z) \sim A \cdot (1 - \mu \cdot z)^{-g-1}$ , we have  $A = a\Gamma(g+1) \approx 0.02$ .

## Improved lower bounds for permutations sortable by 2sip, 2sis and a deque

In this section we derive new lower bounds on the growth rates  $\mu_p$ ,  $\mu_s$  and  $\mu_d$  of permutations sortable by two stacks in parallel, two stacks in series

and a double ended queue, respectively. We deduce lower bounds of 8.18377, 8.18377 and 8.21927 on the growth rates  $\mu_p$ ,  $\mu_s$  and  $\mu_d$ , respectively, which improve on the respective bounds 7.535, 8.156 and 7.890 shown by Albert, Atkinson and Linton [2]. To derive our new bounds, we use the fact that for each of these machines, the class of sortable permutations is sum-closed, that is, if  $\tau$  and  $\pi$  are sortable permutations, then so is  $\tau \oplus \pi$ . It follows that we can write

$$P(t) = \frac{1}{1 - p(t)},$$

where  $p(t)$  is the generating function for primitive 2sip-sortable permutations, and we can write  $D(t)$  and  $S(t)$  in a similar way. Recall that we computed the first  $N = 1336$  coefficients of the generating functions  $P(t)$  of 2sip-sortable permutations. Using these coefficients, we can compute that first  $N$  coefficients of the generating function  $p(t)$  of primitive 2sip-sortable permutations, using the equation above. The degree  $N$  polynomial

$$[t^{\leq N}]p(t) = \sum_{n=0}^N t^n ([t^n]p(t))$$

clearly satisfies  $[t^{\leq N}]p(t) \leq p(t)$ . Hence,

$$P(t) \geq \frac{1}{1 - [t^{\leq N}]p(t)}.$$

For  $N = 1336$ , the radius of convergence of the right hand side of the inequality is  $t_c = 0.122\dots$ , so  $1/t_c = 8.18377\dots$  is a lower bound for the growth rate of 2sip-sortable permutations.

Using exactly the same method, we derive a lower bound of 8.21927... for the growth rate of deque-sortable permutations. We can also apply this method to our 20 term series for 2sis-sortable permutations, however this only yields the lower bound 7.03 on the associated growth rate. Instead, we make the observation that two stacks in series are strictly more powerful than two stacks in parallel, in the sense that any permutation which can be sorted by two stacks in parallel can also be sorted by two stacks in series. This is because two stacks in series can simulate two stacks in parallel. To see this in terms of operation sequences, a 2sip-word can be transformed into a 2sis-word which produces the same permutation by replacing each  $I_1$ ,  $I_2$ ,  $O_1$ ,  $O_2$  with  $\rho$ ,  $\rho\lambda$ ,  $\lambda\mu$ ,  $\mu$  respectively. Hence the lower bound 8.18377 on the growth rate of 2sip-achievable permutations is also a lower bound for the growth rate of 2sis-achievable permutations.

## Chapter 3

# Bounds on the growth rate of 1324-avoiding permutations

### Background on pattern avoiding permutations

A permutation  $\pi$  is said to contain a permutation  $p = a_1a_2 \dots a_k$  if there is some subsequence  $b_1, \dots, b_k$  of  $\pi$  which has the same relative order as  $p$ . That is,  $b_i < b_j$  if and only if  $a_i < a_j$ . In this context, the permutation  $p$  is generally called a pattern. The permutation  $\pi$  is said to *avoid* the pattern  $p$  if it does not contain  $p$ . For example, the permutation 2736514 contains the pattern 4231, since the numbers 7351 have the same relative order as 4231. On the other hand, 2736514 avoids the pattern 1324. The set of permutations  $\pi$  which avoid a pattern  $p$  is denoted  $\text{Av}(p)$ , and the set of those of length  $n$  is denoted  $\text{Av}_n(p)$ . More generally, the set of permutations which avoid all of the patterns  $p_1, p_2, \dots, p_m$  is denoted  $\text{Av}(p_1, p_2, \dots, p_m)$ . The main object of study is the numbers  $|\text{Av}_n(p)|$ , that is, how many permutations of length  $n$  avoid the pattern  $p$ . In 1915, MacMahon [151] solved this for all patterns of length 3, showing that

$$|\text{Av}_n(123)| = |\text{Av}_n(132)| = \frac{1}{n+1} \binom{2n}{n},$$

the  $n$ th Catalan number.

Although permutation patterns were first studied by MacMahon, they became more popular after Knuth studied them in his 1968 book “The Art of Computer Programming” [141]. Knuth’s original motivation was studying permutations sortable by a single stack, which he showed to be exactly the 231-avoiding permutations. Knuth also posed the same question for a variety of more complicated data structures, in particular, those which we

have considered in Chapter 2. Soon after this there were a number of papers on machine sortable permutations [100, 197, 171], but since the 1980s, pattern avoiding permutations have been studied largely for their own sake [183, 37, 109, 195, 206].

For patterns of length 4, there are three distinct enumeration classes or *Wilf classes*. The first class is made up of the permutation patterns 1342, 2413 and their symmetries. The sets  $\text{Av}_n(1342)$  and  $\text{Av}_n(2413)$  were shown to be equinumerous by Stankova [190]. In 1997, Bóna [33] showed that the permutations which avoid either one of these patterns are enumerated by the algebraic counting function

$$\sum_{n=0}^{\infty} |\text{Av}_n(1342)| t^n = \frac{32t}{1 + 20t - 8t^2 - (1 - 8t)^{3/2}}.$$

The next class is made up of the permutation patterns 1234, 1243, 1432, 2143 and their symmetries [191]. In 1990, Gessel enumerated the permutations avoiding 1234 [115], though the following form of the solution was demonstrated by Bousquet-Mélou in 2003 [46]

$$|\text{Av}_n(1234)| = \frac{1}{(n+1)^2(n+2)} \sum_{k=0}^n \binom{2k}{k} \binom{n+1}{k+1} \binom{n+2}{k+1}.$$

The counting function in this case is D-finite, but not algebraic.

The final Wilf class for length 4 patterns has only two representatives: 1324 and 4231, but in these cases the enumeration problem remains unsolved. Numerical work by Conway, Guttmann and Zinn-Justin [71] suggests that the counting function in this case is not D-finite. In this chapter, we present our proof that the growth rate of 1324-avoiding permutations lies between 10.27 and 13.5. This is joint work with David Bevan, Robert Brignall and Jay Pantone.

The exponential *growth rate* of the class  $\text{Av}(\pi)$  is denoted

$$\text{gr}(\text{Av}(\pi)) = \lim_{n \rightarrow \infty} \sqrt[n]{|\text{Av}_n(\pi)|},$$

where  $\text{Av}_n(\pi)$  denotes the set of permutations of length  $n$  that avoid  $\pi$ . This limit is known to exist for any single pattern  $\pi$  as a consequence of the resolution of the Stanley-Wilf conjecture by Marcus and Tardos [154]. In 1999 Arratia conjectured that for a permutation  $\pi$  of length  $k$ , the growth rate  $\text{gr}(\text{Av}(\pi))$  is no more than  $(k-1)^2$  [9], however this was refuted in 2006 by Albert, Elder, Rechnitzer, Westcott and Zabrocki, when they showed that  $\text{gr}(\text{Av}(1324)) > 9.47$  [4]. Recently Fox used a probabilistic argument to show



that  $\text{gr}(\text{Av}(\pi))$  is typically super-polynomial in  $k$  [109], destroying any hope of salvaging Arratia’s conjecture.

For a more thorough introduction to the enumerative theory of permutation classes, see Vatter’s exposition [206]. The topic is also presented in a broader context in books by Bóna [38] and Kitaev [139].

## The growth rate of 1324-avoiding permutations

	Lower	Upper
2004: Bóna [35]		288
2005: Bóna [36]	9	
2006: Albert et al. [4]	9.47	
2012: Claesson, Jelínek and Steingrímsson [65]		16
2014: Bóna [39]		13.93
2015: Bóna [40]		13.74
2015: Bevan [27]	9.81	
<i>This work</i>	10.27	13.5

Table 3.1: A chronology of lower and upper bounds for  $\text{gr}(\text{Av}(1324))$ .

Our interest is in the growth rate of the class  $\text{Av}(1324)$ , the subject of a number of papers over the last decade and a half. For an entertaining essay placing the problem in a wider historical context, see [84]. The history of rigorous lower and upper bounds for  $\text{gr}(\text{Av}(1324))$  is summarised in Table 3.1. In addition to these, Claesson, Jelínek and Steingrímsson [65] make a conjecture regarding the number of 1324-avoiders of each length that have a fixed number of inversions, which if proven would yield an improved upper bound of  $e^{\pi\sqrt{2/3}} \approx 13.002$ .

Conway, Guttmann and Zinn-Justin have computed  $|\text{Av}_n(1324)|$  for all  $n \leq 50$  [70, 71]. By analysing these numbers using differential approximants, they give a numerical estimate for  $\text{gr}(\text{Av}(1324))$  of  $11.600 \pm 0.003$ . They also conjecture that  $|\text{Av}_n(1324)|$  behaves asymptotically as  $A \cdot \mu^n \cdot \lambda^{\sqrt{n}} \cdot n^\alpha$ , for certain estimated constants  $A$ ,  $\mu$ ,  $\lambda$  and  $\alpha$ . If this conjecture were proved, then as a consequence of [113, Theorem 9], it would imply that the counting sequence for 1324-avoiders is not P-recursive (i.e. does not satisfy a linear recurrence with polynomial coefficients), perhaps going some way to explain the difficulties faced in its enumeration.

Our contribution to the investigation of the 1324-avoiders is to establish new rigorous lower and upper bounds on  $\text{gr}(\text{Av}(1324))$ . These rely on a new

structural characterisation of  $\text{Av}(1324)$  as a subclass of an infinite staircase grid class, which we present in the next section. In Section 3.4, we investigate pairs of adjacent cells in the staircase, which we call *dominoes*, and give an exact enumeration (Theorem 3.2). Together with a result concerning *balanced* dominoes, this is sufficient to deduce a new upper bound of 13.5 and a new lower bound of 10.125 on the growth rate of  $\text{Av}(1324)$ , which we present in the following two sections as Theorems 3.8 and 3.9.

The lower bound can be increased by investigating the structure of dominoes in greater detail. In Section 3.7, we prove two asymptotic concentration results, relating to *leaves* and *empty strips*. Section 3.8 then presents a refinement of our staircase construction, a lower bound on the number of ways of combining dominoes, and a technical analysis of the resulting generating function. This yields, in Theorem 3.16, a lower bound on  $\text{gr}(\text{Av}(1324))$  of 10.271.

## Staircase structure

In this section, we present a structural characterisation of  $\text{Av}(1324)$  as a subclass of a larger permutation class. This class is a *staircase class*, which is a special case of an infinite grid class of permutations. We begin by defining finite and infinite grid classes.

Suppose that  $M$  is a  $t \times u$  matrix of (possibly empty) permutation classes, where  $t$  is the number of columns and  $u$  the number of rows. An  $M$ -gridding of a permutation  $\sigma$  of length  $n$  is a pair of sequences  $1 = c_1 \leq \dots \leq c_{t+1} = n + 1$  (the *column dividers*) and  $1 = r_1 \leq \dots \leq r_{u+1} = n + 1$  (the *row dividers*) such that for all  $k \in [t]$  and  $\ell \in [u]$ , the entries of  $\sigma$  whose indices are in  $[c_k, c_{k+1})$  and values in  $[r_\ell, r_{\ell+1})$  are order isomorphic to an element of  $M_{k,\ell}$ . Thus, an  $M$ -gridding of  $\sigma$  partitions the entries of  $\sigma$ , with one part for each cell in  $M$ . A permutation together with one of its  $M$ -griddings is called an  $M$ -gridded permutation.

The *grid class* of  $M$ , denoted  $\text{Grid}(M)$ , consists of all the permutations that have an  $M$ -gridding. We also use  $\text{Grid}^\#(M)$  to denote the set of all  $M$ -gridded permutations, every permutation in  $\text{Grid}(M)$  being present once with each of its  $M$ -griddings.

The definition of a grid class extends naturally for infinite matrices. If  $M$  is an infinite matrix of permutation classes, then the *infinite grid class*  $\text{Grid}(M)$  consists of all the permutations that have an  $M'$ -gridding, for some finite submatrix  $M'$  of  $M$ .

Of direct interest to us are *staircase classes*, infinite grid classes that have a staircase structure (for more on staircase classes, see [5]). Given two



**Proposition 3.1.**  $\text{Av}(1324)$  is contained in the descending  $(\text{Av}(213), \text{Av}(132))$  staircase.

To prove this result, we describe how to construct an explicit gridding of any 1324-avoider in the staircase. Here, and elsewhere in our discussion, we identify a permutation  $\sigma$  with its *plot*, the set of points  $(i, \sigma(i))$  in the Euclidean plane, and refer to its entries as *points*.

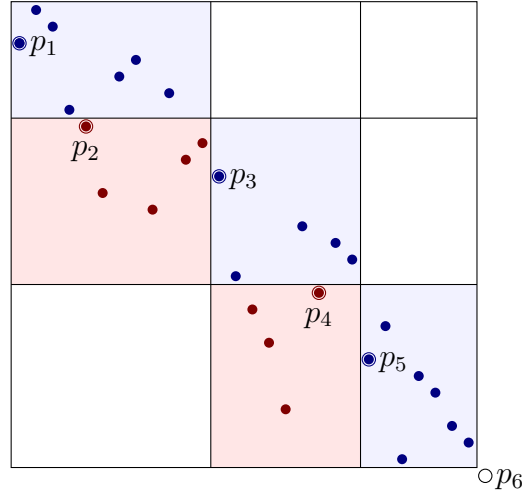


Figure 3.2: The greedy gridding of a 1324-avoider in the staircase.

*Proof.* Consider any  $\sigma \in \text{Av}(1324)$  of length  $n$ . We construct a gridding of  $\sigma$  in the staircase as follows. Let  $p_1$  be the leftmost point of  $\sigma$ , and iteratively identify subsequent points  $p_2, \dots, p_k$  as follows. See Figure 3.2 for an illustration.

- If  $i$  is even, let  $p_i$  be the uppermost point of  $\sigma$  that acts as a 1 in an occurrence of 213 consisting only of points to the right of the column divider adjacent to  $p_{i-1}$ . Insert a row divider immediately above  $p_i$ . If no suitable point exists, terminate.
- If  $i > 1$  is odd, let  $p_i$  be the leftmost point of  $\sigma$  that acts as a 2 in an occurrence of 132 consisting only of points below the row divider adjacent to  $p_{i-1}$ . Insert a column divider immediately to the left of  $p_i$ . If no suitable point exists, terminate.

Since three points are required for an occurrence of 213 or 132, this process terminates after identifying  $k$  points, where  $k \leq \lceil n/2 \rceil$ . Finally, let  $p_{k+1}$  be a virtual point at  $(n+1, 0)$ , below and to the right of all points of  $\sigma$ .

By construction, if  $i \in [2, k+1]$  is even, then the points of  $\sigma$  above  $p_i$  and to the right of the column divider adjacent to  $p_{i-1}$  avoid 213. Analogously, if  $i \in [3, k+1]$  is odd, then the points of  $\sigma$  to the left of  $p_i$  and below the row divider adjacent to  $p_{i-1}$  avoid 132.

Furthermore, if  $i \in [2, k]$  is even, then there are no points of  $\sigma$  below  $p_i$  and to the left of  $p_{i-1}$ , since any such point would form a 1324 with the 213 of which  $p_i$  acts as a 1. Analogously, if  $i \in [3, k]$  is odd, then there are no points of  $\sigma$  to the right of  $p_i$  and above  $p_{i-1}$ , since any such point would form a 1324 with the 132 of which  $p_i$  acts as a 2.

Thus, the column and row dividers specify a valid  $M$ -gridding of  $\sigma$ , where  $M$  is a finite submatrix of the infinite matrix defining the staircase.  $\square$

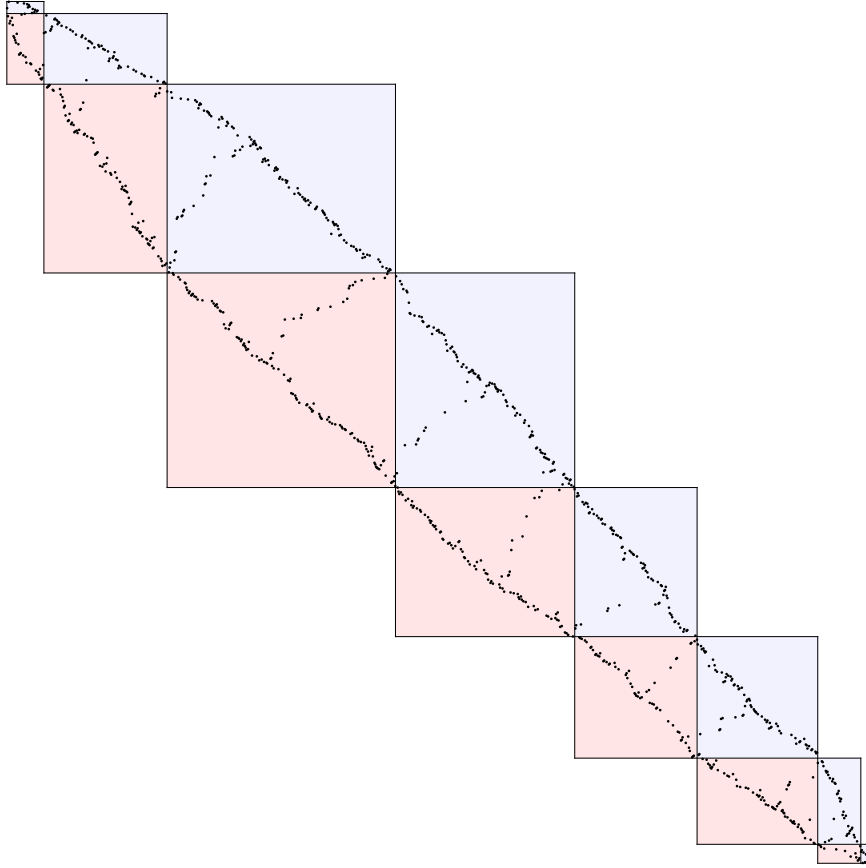


Figure 3.3: The greedy gridding of a 1324-avoider of length 1000.

We call the gridding of a 1324-avoider  $\sigma$  constructed in the proof of Proposition 3.1 the *greedy gridding of  $\sigma$* , because, as we descend the staircase, we

place as many points of  $\sigma$  as possible in each subsequent cell. See Figure 3.3 for an illustration of the greedy gridding of a large permutation.<sup>1</sup>

This structural characterisation has not been presented previously. However, the colouring approach used by Claesson, Jelínek and Steingrímsson in [65] and refined by Bóna in [39, 40] depends on the fact that  $\text{Av}(1324)$  is a subclass of the *merge* of the permutation classes  $\text{Av}(213)$  and  $\text{Av}(132)$ . Given two permutation classes  $\mathcal{C}$  and  $\mathcal{D}$ , their merge, written  $\mathcal{C} \odot \mathcal{D}$ , is the set of all permutations whose entries can be coloured blue and red so that the blue subsequence is order isomorphic to a member of  $\mathcal{C}$  and the red subsequence is order isomorphic to a member of  $\mathcal{D}$ .

The descending staircase is contained in the merge  $\text{Av}(213) \odot \text{Av}(132)$ , since points gridded in the upper,  $\text{Av}(213)$ , cells collectively avoid 213, and the remaining points gridded in the lower,  $\text{Av}(132)$ , cells collectively avoid 132. Thus our new characterisation is a refinement of that used previously. However, the growth rate of the staircase and that of the merge are both 16 (see [5]), so Proposition 3.1 doesn't immediately yield any improvement over the upper bound in [65].

## 1324-avoiding dominoes

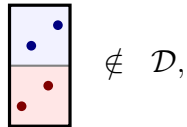


Figure 3.4: Four distinct small dominoes.

To establish bounds on the growth rate of  $\text{Av}(1324)$ , we investigate pairs of adjacent cells in the griddings of 1324-avoiders in the staircase. We define a 1324-avoiding vertical *domino* to be a two-cell *gridded permutation* in  $\text{Grid}^{\#} \left( \begin{smallmatrix} \text{Av}(213) \\ \text{Av}(132) \end{smallmatrix} \right)$  whose underlying permutation avoids 1324. See Figure 3.4 for an illustration of four dominoes, the two at the left being distinct griddings of 34251, and the two at the right being distinct griddings of 31524.

<sup>1</sup>The data for Figure 3.3 was provided by Einar Steingrímsson from the investigations he describes in [195, Footnote 4].

Let  $\mathcal{D}$  be the set of dominoes. It is important to note that



since  $\mathcal{D}$  consists of gridded 1324-avoiders. Moreover, within the grid class  $\text{Grid}\left(\begin{smallmatrix} \text{Av}(213) \\ \text{Av}(132) \end{smallmatrix}\right)$ , this is the *only* arrangement of points that must be avoided, since it is the only possible gridding of 1324 in the two cells. With the cell divider in any other position, either the top cell contains a 213 or the bottom cell contains a 132.

In this section we enumerate the gridded permutations in  $\mathcal{D}$  by placing them in bijection with certain arch configurations, proving the following theorem.

**Theorem 3.2.** *The number of  $n$ -point dominoes is  $\frac{2(3n+3)!}{(n+2)!(2n+3)!}$ . Consequently,  $\text{gr}(\mathcal{D}) = 27/4$ .*

This theorem, along with the result that *balanced* dominoes have the same growth rate (Proposition 3.7), gives us enough information to calculate improved upper and lower bounds for the growth rate of  $\text{Av}(1324)$ .

In order to prove Theorem 3.2, our first task is to establish a functional equation for the set of dominoes  $\mathcal{D}$ . We do this by representing dominoes as configurations consisting of an interleaved pair of arch systems, one for each of the two cells.

## Arch systems

Let an  $n$ -point *arch system* consist of  $n$  points on a horizontal line together with zero or more noncrossing arcs, all on the same side of the line, connecting distinct pairs of points, such that no point is the left endpoint of more than one arc and no point is the right endpoint of more than one arc. See Figure 3.5. Note that these are not non-crossing *matchings*.

These arch systems are equinumerous with domino cells.<sup>2</sup> We make use of a bijection in which arcs correspond to occurrences of 12 in the cells, having the form  $k(k+1)$  for some value  $k$ .

**Proposition 3.3.** *Both  $\text{Av}_n(213)$  and  $\text{Av}_n(132)$  are in bijection with  $n$ -point arch systems.*

---

<sup>2</sup>Despite being enumerated by the Catalan numbers, these specific arch systems are, rather surprisingly, not included in Stanley's book [194].

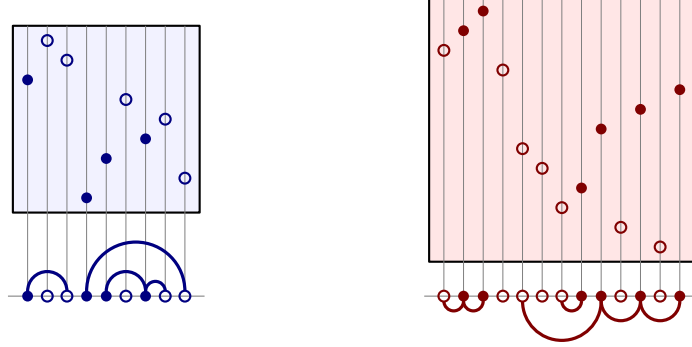


Figure 3.5: A 213-avoider and a 132-avoider with their arch systems.

*Proof.* We define a mapping  $\Lambda$  from  $\text{Av}(213)$  and  $\text{Av}(132)$  to arch systems. This mapping is illustrated in Figure 3.5. Given a 213-avoiding or 132-avoiding permutation  $\sigma$  of length  $n$ , let the points of the corresponding arch system  $\Lambda(\sigma)$  be positioned at  $1, \dots, n$  on the line. For each pair  $i, j$  with  $1 \leq i < j \leq n$ , connect the points at  $i$  and  $j$  with an arc if and only if  $\sigma(j) = \sigma(i) + 1$ .

The result is a valid arch system. Crossing arcs could only result from an occurrence in  $\sigma$  of either 1324 or 3142, both of which contain both 213 and 132, and by construction no point can be the left endpoint of more than one arc or the right endpoint of more than one arc.

In the converse direction, we recursively define mappings  $\Pi_{213}$  and  $\Pi_{132}$  from arch systems to  $\text{Av}(213)$  and  $\text{Av}(132)$  respectively, such that for any arch system  $\alpha$ , we have

$$\Lambda(\Pi_{213}(\alpha)) = \Lambda(\Pi_{132}(\alpha)) = \alpha. \quad (3.1)$$

Trivially, in both cases, we map the 0-point arch system to the empty permutation and the 1-point arch system to the singleton permutation 1.

Now, suppose  $\alpha$  is the concatenation  $\alpha_1\alpha_2$  of two nonempty arch systems. Then  $\Pi_{213}(\alpha)$  is the skew sum  $\Pi_{213}(\alpha_1) \ominus \Pi_{213}(\alpha_2)$ , a copy of  $\Pi_{213}(\alpha_1)$  being positioned to the upper left of  $\Pi_{213}(\alpha_2)$ .  $\Pi_{132}(\alpha)$  is similar. Otherwise,  $\Lambda(\Pi_{213}(\alpha))$  and  $\Lambda(\Pi_{132}(\alpha))$  would have an arc connecting some point of  $\alpha_1$  to some point of  $\alpha_2$ .

Finally, suppose  $\alpha$  is a sequence of  $k$  (possibly empty) arch systems,  $\alpha_1, \dots, \alpha_k$ , enclosed in  $k$  connected arcs, like  $\underbrace{\alpha_1 \dots \alpha_k}_{\text{arc}}$ . Then  $\Pi_{213}(\alpha)$  consists of  $\Pi_{213}(\alpha_1 \dots \alpha_k)$  above the increasing permutation  $12 \dots (k+1)$ , where  $\Pi_{213}(\alpha_i)$  is between  $i$  and  $i+1$  for each  $i$ . See Figure 3.6 for an illustration. To satisfy (3.1), the endpoints of the arcs must map to consecutive increasing values in the permutation, and each  $\Pi_{213}(\alpha_i)$  must be



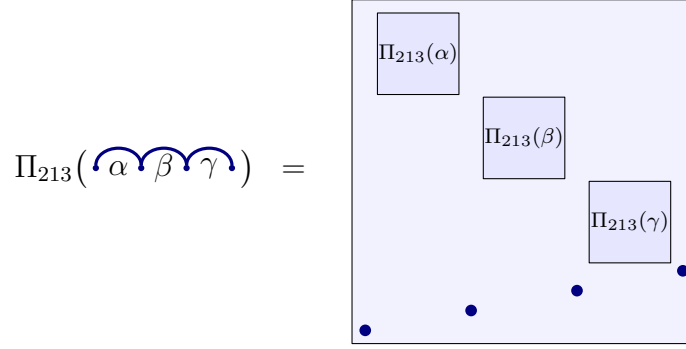


Figure 3.6: Mapping an arch system to a 213-avoider.

above  $\Pi_{213}(\alpha_{i+1})$ . To avoid creating an occurrence of 213, each nonempty  $\Pi_{213}(\alpha_i)$  must be above  $i$  and  $i + 1$ . Analogously, to avoid creating a 132,  $\Pi_{132}(\alpha)$  consists of  $\Pi_{132}(\alpha_1 \dots \alpha_k)$  *below* an increasing permutation of length  $k + 1$ .  $\square$

As an aside, we note that the proof of Proposition 3.3 can easily be adapted to establish that in  $\text{Av}_n(213)$  and  $\text{Av}_n(132)$  each permutation is uniquely determined by the set consisting of the pairs of values comprising its ascents.

## Arch configurations

A domino is comprised of a 213-avoiding top cell and a 132-avoiding bottom cell. Thus, by Proposition 3.3, corresponding to each domino is an *arch configuration* consisting of an interleaved pair of arch systems. See Figure 3.7 for an illustration. In the figures, the arch system for the top cell is shown above the line, and that for the bottom cell is below the line. Isolated points are marked with a short strut to indicate to which arch system they belong.

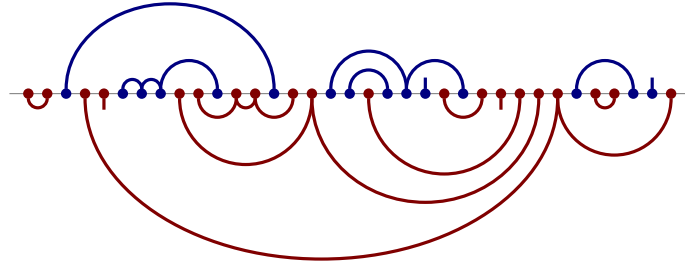
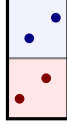


Figure 3.7: The arch configuration for a domino.

Recall that the only restriction on the cells in a domino is that the following arrangement of points (forming a 1324) must be avoided.



The arch configuration corresponding to this is  $\curvearrowright\curvearrowleft$ . Indeed, avoiding this pattern of arcs in an arch configuration is equivalent to avoiding 1324 in a domino.

**Proposition 3.4.** *The set  $\mathcal{D}$  of dominoes is in bijection with arch configurations that do not contain the pattern  $\curvearrowright\curvearrowleft$ .*

*Proof.* By the bijection used in the proof of Proposition 3.3, an arch configuration contains an occurrence of  $\curvearrowright\curvearrowleft$  if and only if the corresponding pair of cells contains an occurrence of 1324 of the form  $k\ell(k+1)(\ell+1)$ , for values  $k$  and  $\ell$  such that  $\ell > k+1$ . So, if an arch configuration contains  $\curvearrowright\curvearrowleft$ , the corresponding gridded permutation contains 1324.

For the converse, it suffices to show that if a permutation gridded in  $\text{Grid}\left(\begin{smallmatrix} \text{Av}(213) \\ \text{Av}(132) \end{smallmatrix}\right)$  contains an occurrence of 1324, then it contains some, possibly distinct, occurrence of 1324 that has the form  $k\ell(k+1)(\ell+1)$ . Suppose  $acbd$  is an occurrence of 1324, gridded in  $\text{Grid}\left(\begin{smallmatrix} \text{Av}(213) \\ \text{Av}(132) \end{smallmatrix}\right)$ , where  $a < b < c < d$ . Then  $a$  and  $b$  are in the bottom, 132-avoiding, cell. Consider the set of values in the interval  $I = \{a, a+1, \dots, b-1\}$ . These must all occur to the left of  $b$ , otherwise a 132 would be formed. Let  $a+i$ , where  $i \geq 0$ , be the greatest element of  $I$  that occurs to the left of  $c$ ; this value must exist since  $a$  itself occurs before  $c$ . Then  $(a+i)c(a+i+1)d$  is an occurrence of 1324 in which the first and third values differ by one.

Applying an analogous argument to the interval  $J = \{c+1, \dots, d-1, d\}$  then yields  $j \geq 0$  such that  $(a+i)(d-j-1)(a+i+1)(d-j)$  is an occurrence of 1324 with the required form.  $\square$

To enumerate dominoes, we construct a functional equation for arch configurations, which we then solve. We build arch configurations from left to right. A vertical line positioned between two points of an arch configuration may intersect some arcs. We call the partial arch configuration to the left of such a line an *arch prefix*; any arcs intersected by the line are *open*.

Let  $\mathcal{A}$  be the set of arch prefixes with no open *upper* arcs, and let  $A(v) = A(z, v)$  be the ordinary generating function for  $\mathcal{A}$ , in which  $z$  marks points and  $v$  marks open *lower* arcs. Thus,  $A(0) = A(z, 0)$  is the generating function for the set of dominoes  $\mathcal{D}$ .

**Proposition 3.5.** *The generating function  $A(v) = A(z, v)$ , for the set  $\mathcal{A}$  of arch prefixes with no open upper arcs, in which  $z$  marks points and  $v$  marks open lower arcs, satisfies the functional equation*

$$A(v) = \frac{1}{1 - zA(v)} + z(1 + v) \left( A(v) + \frac{A(v) - A(0)}{v} \right). \quad (3.2)$$

*Proof.* There are six possible ways in which a non-empty element of  $\mathcal{A}$  can be decomposed, depending on its rightmost point. These are illustrated in Figure 3.8.

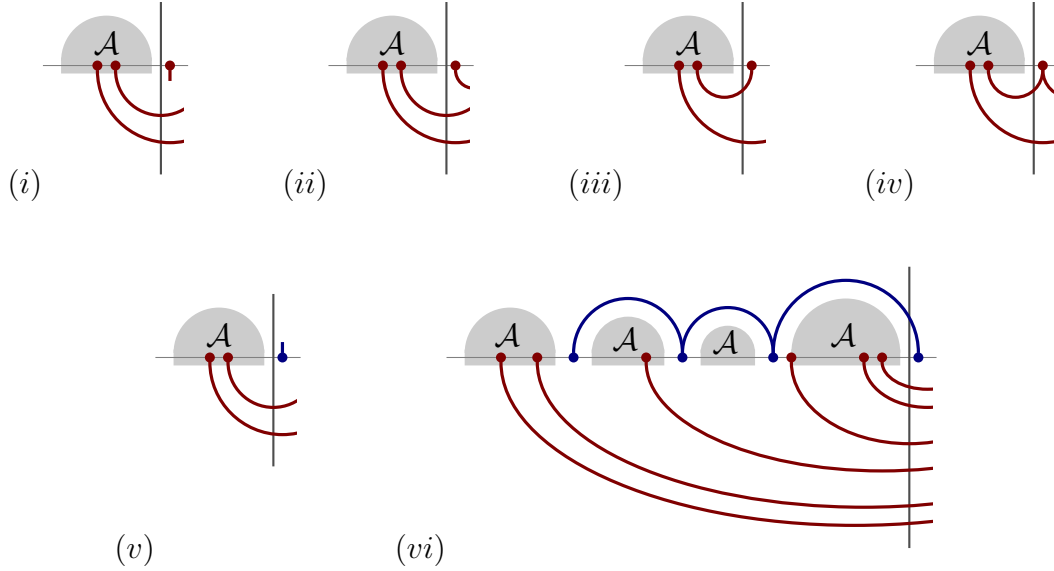


Figure 3.8: The six ways of decomposing a non-empty arch prefix in  $\mathcal{A}$ .

If the rightmost point belongs to the lower arch system, then there are four cases: (i) an isolated point, (ii) the left endpoint of an arc, (iii) the right endpoint of an arc, and (iv) both the left and right endpoint of an arc. These contribute the following terms to the functional equation for  $A(v)$ :

$$(i) \quad zA(v) \quad (ii) \quad zvA(v) \quad (iii) \quad zv^{-1}(A(v) - A(0)) \quad (iv) \quad z(A(v) - A(0)).$$

If the rightmost point belongs to the upper arch system, then, since there are no open upper arcs, it is either (v) an isolated point, or else (vi) the right endpoint of an arc. In the former case, this contributes  $zA(v)$  to the functional equation for  $A(v)$ . In the latter case, the arch prefix decomposes into a connected sequence of one or more upper arcs, each enclosing an

element of  $\mathcal{A}$  (possibly empty), preceded by a further initial element of  $\mathcal{A}$  (also possibly empty). This makes a contribution of

$$\frac{z^2 A(v)^2}{1 - z A(v)}$$

to the functional equation for  $A(v)$ .

Summing these terms, including a term for the empty prefix, and simplifying, yields the functional equation in the statement of the proposition.  $\square$

## The enumeration of dominoes

To complete the proof of Theorem 3.2, we employ resultant methods to eliminate the variables  $v$  and  $A(v)$  from the functional equation (3.2). This yields a minimal polynomial for  $A(0)$  which we then use to derive the closed-form formula for the number of dominoes and their exponential growth rate.

*Proof of Theorem 3.2.* Clearing denominators from (3.2) and moving all terms to one side yields

$$0 = P(A(v), A(0), z, v)$$

where  $P$  is the polynomial

$$P(x, y, z, v) = (zv - z^2(1+v)^2)x^2 + z^2(1+v)xy + (z(1+v)^2 - v)x - z(1+v)y + v.$$

The presence of the term  $x^2$  indicates that the kernel method does not apply here. Instead, we use a more general method of Bousquet-Mélou and Jehanne [51] which says that  $A(v)$  and  $v$  can be eliminated from the functional equation via *iterated discriminants*. Specifically, define

$$Q(y, z) = \text{discrim}_v(\text{discrim}_x(P(x, y, z, v))).$$

Then it follows that the minimal polynomial for  $A(0)$  is one of the irreducible factors of  $Q(y, z)$ . Performing the calculation, we find that

$$Q(y, z) = -256z^8 R_1(y, z)^2 R_2(y, z),$$

where

$$\begin{aligned} R_1(y, z) &= z^3 y^2 + z(1 - 4z)y + 4z - 1, \\ R_2(y, z) &= z^4 y^3 + 2z^2(3z + 1)y^2 + (12z^2 - 10z + 1)y + 8z - 1. \end{aligned}$$

The two series solutions of  $0 = R_1(y, z)$  begin  $y = z^{-1} + O(1)$  and  $y = -z^{-2} + O(z^{-1})$ , which do not match the known initial terms of  $A(0)$ . Therefore, it is  $R_2$  that is a minimal polynomial for  $A(0)$ .

We verify that, for each  $n$ , the coefficient of  $z^n$  in the series expansion of  $A(0)$  is given by

$$\frac{2(3n+3)!}{(n+2)!(2n+3)!},$$

by using *Mathematica* [212].

```
minpoly[y_] := z^4 y^3 + 2z^2(3z + 1)y^2 + (12z^2 - 10z + 1)y + 8z - 1
series = Sum[2(3n + 3)!/((n + 2)!(2n + 3)!) z^n, {n, 0, Infinity}]
(2(-1 - 3z + Hypergeometric2F1[-2/3, -1/3, 1/2, 27z/4]))/(3z^2)

minpoly[series] // FunctionExpand // Simplify
0
```

The first command assigns the known minimal polynomial for  $A(0)$  to the variable `minpoly`. The second command creates the power series that we want to verify is equal to  $A(0)$ ; *Mathematica* deduces a nice form for this. The final command substitutes the power series into the minimal polynomial and simplifies. The result is 0, so the power series satisfies the minimal polynomial. Since the initial terms of the power series coincide with those of  $A(0)$  and not with those of the other roots of  $R_2$ , this completes the proof of the first part of Theorem 3.2.

To derive the growth rate, note that the exponential growth rate of an algebraic generating function (and, in fact, a complete asymptotic expansion) can be derived from the minimal polynomial using the method outlined by Flajolet and Sedgewick [106, Note VII.36]. The exponential growth rate must be the reciprocal of one of the roots of the discriminant of the minimal polynomial with respect to  $y$ . Since

$$\text{discrim}_y(z^4 y^3 + 2z^2(3z + 1)y^2 + (12z^2 - 10z + 1)y + 8z - 1) = -z^5(27z - 4)^3,$$

and with the knowledge that algebraic generating functions for combinatorial sequences are analytic at the origin [140, Proposition 3.1], we conclude that the exponential growth rate for the power series of  $A(0)$  is  $27/4 = 6.75$ .  $\square$

The counting sequence for dominoes is [A000139](#) in *OEIS* [198]. Among other things, this enumerates West-two-stack-sortable permutations [213], rooted nonseparable planar maps [57] and a class of branching polyominoes known as *fighting fish* [77, 78, 101]. So far, we have not been able to establish a bijection between dominoes and any of these structures.

**Problem 3.6.** Find a bijection between 1324-avoiding dominoes and another combinatorial class known to be equinumerous.

## Balanced dominoes

We say that a domino is *balanced* if its top cell contains the same number of points as its bottom cell. Let  $\mathcal{B}$  be the set of balanced dominoes and  $\mathcal{B}_m$  be the set of balanced dominoes having a total of  $2m$  points,  $m$  points in each cell. We define the growth rate of balanced dominoes to be  $\text{gr}(\mathcal{B}) = \lim_{m \rightarrow \infty} \sqrt[2m]{|\mathcal{B}_m|}$ . We prove that the growth rate of *balanced* dominoes is the same as that of *all* dominoes.

**Proposition 3.7.** *The growth rate of balanced dominoes is  $27/4$ .*

In the proof, we use two elementary manipulations of dominoes. Given a domino  $\sigma$ , let the  $180^\circ$  rotation of  $\sigma$  be denoted  $\hat{\sigma}$ . This is itself a valid domino. Also, given two dominoes  $\sigma$  and  $\tau$ , define  $\sigma \otimes \tau$  to be the domino whose arch configuration is produced by concatenating the arch configurations of  $\sigma$  and  $\tau$ .

*Proof of Proposition 3.7.* Let  $d(t, b)$  denote the number of  $(t+b)$ -point dominoes with  $t$  points in the top cell and  $b$  points in the bottom cell. For a given  $m$ , let  $t_m$  be a value of  $t$  that maximises  $d(t, m-t)$ . Let  $d_{\max} = d(t_m, m-t_m)$  be this maximal value. Since  $0 \leq t \leq m$ , there are only  $m+1$  possible choices for  $t_m$ . Hence by the pigeonhole principle,

$$d_{\max} \geq \frac{|\mathcal{D}_m|}{m+1}.$$

Let  $\sigma$  and  $\tau$  be any two  $m$ -point dominoes with  $t_m$  points in the top cell and  $m-t_m$  points in the bottom cell. Consider the domino  $\rho = \sigma \otimes \hat{\tau}$ , whose arch configuration is constructed by concatenating the arch configuration of  $\sigma$  and the arch configuration of the  $180^\circ$  rotation of  $\tau$ . This is a balanced domino in  $\mathcal{B}_m$ . Moreover,  $\sigma$  and  $\tau$  can be recovered from  $\rho$  simply by splitting its arch configuration into two halves. Thus,

$$|\mathcal{B}_m| \geq d_{\max}^2 \geq \frac{|\mathcal{D}_m|^2}{(m+1)^2}.$$

Since it is also the case that  $|\mathcal{D}_{2m}| \geq |\mathcal{B}_m|$ , it follows, by taking the  $2m$ th root, and the limit as  $m$  tends to infinity, that  $\text{gr}(\mathcal{B}) = \text{gr}(\mathcal{D}) = 27/4$ .  $\square$

## An upper bound

In this section, we use the results of Section 3.4 to establish a new upper bound on the growth rate of the 1324-avoiders. Our upper bound follows from the fact that we can split a 1324-avoider, gridded in the staircase, in such a way as to produce a domino.

**Theorem 3.8.** *The growth rate of  $\text{Av}(1324)$  is at most  $27/2 = 13.5$ .*

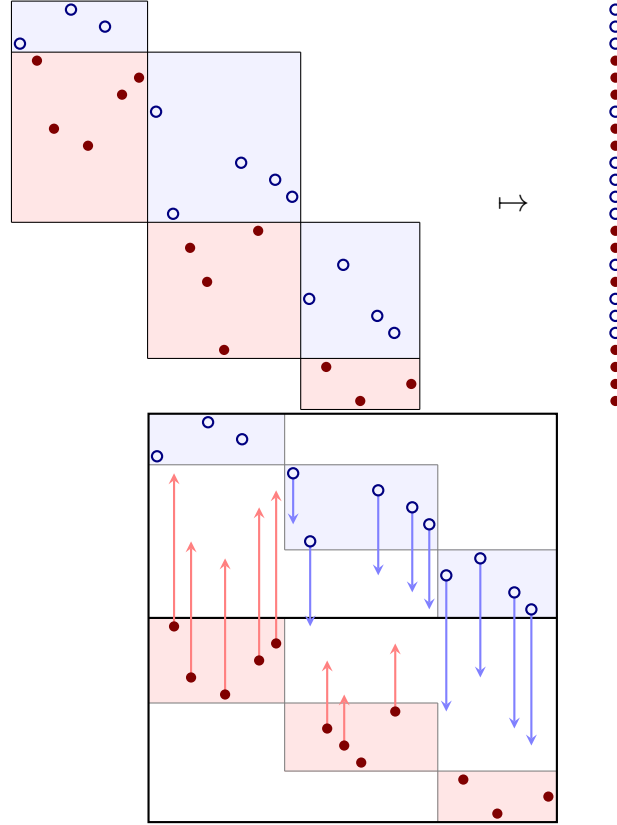


Figure 3.9: Mapping a greedy-gridded 1324-avoider to a binary word and a domino.

*Proof.* We define an injection from  $\text{Av}_n(1324)$  into the Cartesian product  $\{\circ, \bullet\}^n \times \mathcal{D}_n$ , for every  $n \geq 1$ , each permutation being mapped to a pair consisting of a binary word (over the alphabet  $\{\circ, \bullet\}$ ) and a domino. See Figure 3.9 for an illustration. Given a 1324-avoider  $\sigma$ , let  $\sigma^\#$  be the greedy gridding of  $\sigma$  in the descending  $(\text{Av}(213), \text{Av}(132))$  staircase.

The binary word is constructed by reading the points of  $\sigma$  from top to bottom and recording a ring ( $\circ$ ) if the point is in an upper,  $\text{Av}(213)$ , cell of  $\sigma^\#$ , and recording a disk ( $\bullet$ ) if it is in a lower,  $\text{Av}(132)$ , cell.

The domino is constructed by placing all the points from the upper cells of  $\sigma^\#$  in the top cell of the domino, retaining their horizontal positions, and similarly placing the points from the lower cells of  $\sigma^\#$  in the bottom cell of the domino. The result is a valid domino since the points gridded in the upper cells of  $\sigma^\#$  collectively avoid 213, the points gridded in the lower cells

collectively avoid 132 and no additional occurrence of 1324 can be created by splitting  $\sigma^\#$  in this way.

This mapping is an injection, because the original permutation  $\sigma$  can be recovered from the domino by repositioning the points vertically according to the information in the binary word, as illustrated by the arrows in Figure 3.9.

Since there are  $2^n$  binary words of length  $n$  and, by Theorem 3.2, the growth rate of the set of dominoes  $\mathcal{D}$  is  $27/4$ , the union of the Cartesian products has growth rate  $2 \times 27/4 = 13.5$ . The existence of the injection establishes that this value is an upper bound on the growth rate of  $\text{Av}(1324)$ .  $\square$

The use of an arbitrary binary word to record the vertical interleaving of the points is very rudimentary. One would hope that the approach could be refined by recording this information as decorations on the domino in such a way as to yield a tighter upper bound, but we have not been able to do so.

## An initial lower bound

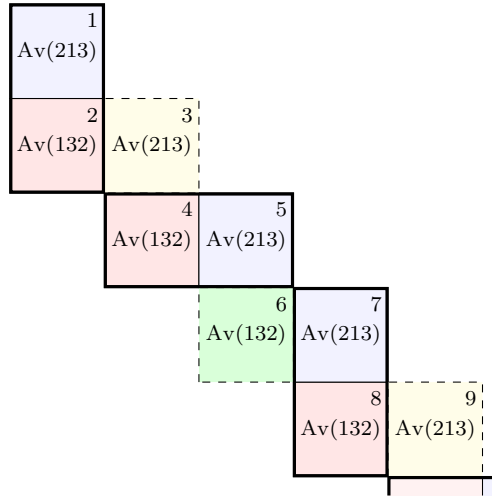


Figure 3.10: The decomposition of the staircase into dominoes and connecting cells.

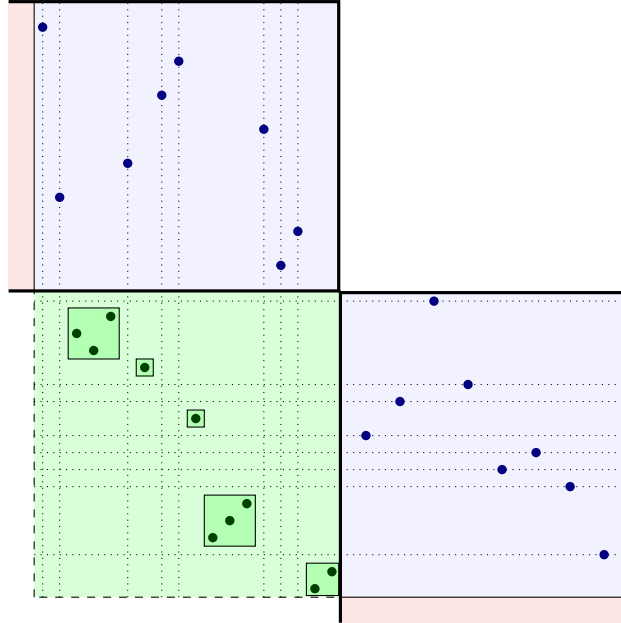
Our lower bounds depend on exploiting a specific partitioning of the staircase. We decompose the staircase into an alternating sequence of dominoes and individual *connecting cells*. See Figure 3.10 for an illustration. In the figure, dominoes are bordered by thick black lines and connecting cells have



dashed borders. Specifically, if we number the cells  $1, 2, \dots$ , descending from the top left, as in the figure, then the decomposition is as follows. For each  $j \geq 0$ :

- Cells numbered  $6j + 1$  and  $6j + 2$  form a (vertical) domino.
- Cells numbered  $6j + 3$  are connecting cells avoiding 213.
- Cells numbered  $6j + 4$  and  $6j + 5$  form a domino reflected about the line  $y = x$  (a *horizontal domino*). The left cell avoids 132 and the right cell avoids 213.
- Cells numbered  $6j + 6$  are connecting cells avoiding 132.

Observe that any occurrence of 1324 in the staircase is contained in a pair of adjacent cells, with two points in each cell. By definition, dominoes avoid 1324. So, to avoid 1324 in this decomposition of the staircase, it is only necessary to guarantee that an occurrence of 1324 is not created from two points in a connecting cell and two points in an adjacent domino cell.



*Figure 3.11:* Interleaving the skew indecomposable components in a connecting cell with the points in the two adjacent domino cells.

Recall that every permutation has a unique representation as the skew sum of a sequence of skew indecomposable components. For brevity, we will refer to a skew indecomposable component simply as a *component*. To ensure that there is no occurrence of 1324, it is sufficient to require that

every point in a domino cell is positioned *between* the components in the adjacent connecting cells. For example, if a domino cell is to the right of a connecting cell, then this restriction ensures that there is no occurrence of 132 in which the 13 is in the connecting cell and the 2 is in the domino cell. See Figure 3.11 for an illustration of a 132-avoiding connecting cell and its adjacent domino cells.

This construction enables us to establish a new lower bound on the growth rate of 1324-avoiders.

**Theorem 3.9.** *The growth rate of  $\text{Av}(1324)$  is at least  $81/8 = 10.125$ .*

To prove this, we take an approach similar to that used by Bevan in [27].

*Proof.* For each  $k \geq 1$ , let  $\mathcal{P}_k$  be the set of gridded permutations, gridded in the first  $3k$  cells of the staircase, decomposed as described above, with every point in a domino cell positioned between the skew indecomposable components in adjacent connecting cells, satisfying the following three conditions.

- Each domino cell contains  $14k$  points.
- Each connecting cell contains  $8k$  points.
- The permutation in each connecting cell has  $7k$  skew indecomposable components.

Each element of  $\mathcal{P}_k$  is a gridded  $36k^2$ -point permutation. The number of these gridded permutations is exactly

$$|\mathcal{P}_k| = |\mathcal{B}_{14k}|^k |\mathcal{C}_{8k, 7k}|^k \binom{21k}{14k}^{2k-1},$$

where  $\mathcal{B}_n$  is, as before, the set of balanced dominoes with  $n$  points in each cell, and  $\mathcal{C}_{n,c}$  is the set of  $n$ -point 213-avoiders (or 132-avoiders) with  $c$  skew indecomposable components. The final binomial coefficient counts the number of possible ways of interleaving  $14k$  points in a domino cell with  $7k$  skew indecomposable components in an adjacent connecting cell.

From Proposition 3.7, we know that  $|\mathcal{B}_n| = (27/4)^{2n} \cdot \theta(n)$ , where the function  $\theta$  satisfies  $\lim_{n \rightarrow \infty} \sqrt[n]{\theta(n)} = 1$ . It is also known that  $|\mathcal{C}_{n,c}| = \frac{c}{n} \binom{2n-c-1}{n-1}$ , since  $\mathcal{C}_{n,c}$  is equinumerous with the number of  $n$ -vertex Catalan forests with  $c$  trees (see [106] Example III.8).

Thus, using Stirling's approximation to determine the asymptotics of the binomial coefficients,

$$\begin{aligned} \lim_{k \rightarrow \infty} |\mathcal{P}_k|^{1/36k^2} &= \lim_{k \rightarrow \infty} \left[ \left( \frac{27}{4} \right)^{28k^2} \theta(14k)^k \cdot \left( \frac{7}{8} \right)^k \binom{9k-1}{8k-1}^k \cdot \binom{21k}{14k}^{2k-1} \right]^{1/36k^2} \\ &= \frac{3^{7/3}}{4^{7/9}} \cdot \frac{3^{1/2}}{2^{2/3}} \cdot \frac{3^{7/6}}{2^{7/9}} = \frac{81}{8}. \end{aligned}$$

Since any  $n$ -point permutation can be gridded in  $j$  cells in at most  $\binom{n+j-1}{j-1}$  ways, the number of ways of gridding a  $36k^2$ -point permutation in  $3k$  cells is no more than  $(6k)^{6k}$ . Hence,

$$|\text{Av}_{36k^2}(1324)| \geq |\mathcal{P}_k| \cdot (6k)^{-6k},$$

and thus  $81/8$  is a lower bound on the growth rate of  $\text{Av}(1324)$ .  $\square$

## Domino substructure

To improve the lower bound of Theorem 3.9, we investigate the structure of dominoes in greater detail. Specifically we prove two concentration results. We say that a sequence of random variables  $X_1, X_2, \dots$  is *asymptotically concentrated at  $\mu$*  if, for any  $\varepsilon > 0$ , for all sufficiently large  $n$ ,

$$\mathbb{P}[|X_n - \mu| \leq \varepsilon] > 1 - \varepsilon.$$

We consider two substructures, which we call *leaves* and *empty strips*, definitions of which are given below. For both, we determine the expected number in an  $n$ -point domino cell and establish that their proportion is concentrated at its mean. As a consequence, almost all dominoes contain “many” leaves and “many” empty strips. Thus, when we refine our staircase construction in the next section, we make use of dominoes that have lots of leaves and lots of empty strips.

### Leaves

Recall that the right-to-left maxima of a permutation are those entries having no larger entry to the right. Similarly, left-to-right minima are those entries having no smaller entry to the left. We say that a point in the top, 213-avoiding, cell of a domino is a *leaf* if it is a right-to-left maximum of the permutation. Analogously, a point in the bottom, 132-avoiding, cell of a

domino is a leaf if it is a left-to-right minimum of the permutation. (These correspond to leaves of the acyclic *Hasse graphs* of the cells; see [27, 47].) In Figure 3.5 on page 81, the leaves are shown as rings.

Recall, from Proposition 3.3, our bijection between domino cells and arch systems. Under this bijection, leaves in a 213-avoiding cell correspond exactly to points which are not the left ends of arcs, and leaves in a 132-avoiding cell correspond to points which are not the right ends of arcs (see Figure 3.5). Thus, adapting Proposition 3.5, if  $A(v, t) = A(z, v, t)$  satisfies the functional equation

$$A(v, t) = 1 + \frac{ztA(v, t)}{1 - zA(v, t)} + z(1 + v) \left( A(v, t) + \frac{A(v, t) - A(0, t)}{v} \right),$$

then  $A(0, t) = A(z, 0, t)$  is the bivariate generating function for dominoes in which  $z$  marks points and  $t$  marks leaves in the top cell.

We want to know how many leaves we can expect to find in a domino cell. We calculate the expected number explicitly.

**Proposition 3.10.** *The total number of leaves in the top cells of all  $n$ -point dominoes is*

$$\frac{5(3n + 1)!}{(n - 1)!(2n + 3)!}.$$

*Consequently, the expected number of leaves in an  $n$ -point domino is asymptotically  $5n/9$ .*

In this and subsequent proofs, we use  $\partial_x f$  to denote the partial derivative  $\partial f / \partial x$ .

*Proof.* The total number of leaves in the top cells of all  $n$ -point dominoes is given by the coefficient of  $z^n$  in  $\partial_t A(0, t)|_{t=1}$ . To calculate this, we use the same technique as in the proof of Theorem 3.2, finding a minimal polynomial  $P_1(y, z, t)$  of degree 7 in  $y$  for  $A(0, t)$ , that is too long to display here.

Differentiating the equation  $0 = P_1(y, z, t)$  with respect to  $t$  yields  $0 = P_2(y, \partial_t y, z, t)$ , where  $P_2$  is a polynomial. We wish now to eliminate  $y$  from  $P_2$  so that a minimal polynomial for  $\partial_t A(0, t)$  remains. This is achieved by computing the resultant of  $P_1$  and  $P_2$  with respect to their first arguments. We find that

$$\text{Res}(P_1(y, z, t), P_2(y, y_1, z, t), y) = Q(z, t)R(y_1, z, t),$$

where  $Q(z, t)$  is a polynomial only in  $z$  and  $t$ , and  $R$  is irreducible.

We conclude therefore that  $R(y, z, t)$  is a minimal polynomial for  $\partial_t A(0, t)$ . Substituting  $t = 1$  shows that  $R(y, z, 1)$  factors into two terms, one of which

must be a minimal polynomial for  $\partial_t A(0, t)|_{t=1}$ . By computing initial terms in the power series expansion of the roots of each factor, we deduce that  $\partial_t A(0, t)|_{t=1}$  is a root of

$$z^3 y^3 + 5z^2 y^2 + (5z - 1)y + z.$$

It can be verified that the coefficient of  $z^n$  in the power series expansion of  $\partial_t A(0, t)|_{t=1}$  is

$$\frac{5(3n+1)!}{(n-1)!(2n+3)!},$$

using *Mathematica* as in the proof of Theorem 3.2, or otherwise. Therefore, the expected number of leaves in the top cell of a domino with  $n$  points is

$$\frac{\frac{5(3n+1)!}{(n-1)!(2n+3)!}}{\frac{2(3n+3)!}{(n+2)!(2n+3)!}} = \frac{5n(n+2)}{6(3n+2)},$$

from which it follows by symmetry that the expected number of leaves in an  $n$ -point domino is asymptotically  $5n/9$ .  $\square$

The sequence of coefficients of the power series for  $\partial_t A(0, t)|_{t=1}$  is [A102893](#) in *OEIS* [198]. This has been shown by Noy [163] to count the number of noncrossing trees on a circle with  $n+1$  edges and root degree at least 2. It would be interesting to find a bijection between these objects and the leaves of 1324-avoiding dominoes.

We need to show that the proportion of points that are leaves is asymptotically concentrated. We calculate the variance directly.

**Proposition 3.11.** *The proportion of leaves in the top cell of an  $n$ -point domino is asymptotically concentrated at its mean.*

*Proof.* Let  $E_n$  be the expected number of leaves in the top cell of an  $n$ -point domino, given by Proposition 3.10, and let  $V_n$  be the variance of the number of leaves in the top cell of an  $n$ -point domino. As described in Flajolet and Sedgewick [106, Proposition III.2],

$$V_n = \frac{[z^n] \partial_{tt} A(0, t)|_{t=1}}{[z^n] A(0, 1)} + E_n - E_n^2.$$

We start by determining  $\partial_{tt} A(0, t)|_{t=1}$ . The minimal polynomial for  $\partial_{tt} A(0, t)|_{t=1}$  is computed from the minimal polynomial for  $\partial_t A(0, t)$  using the same method

as in the proof of Proposition 3.10. One finds that  $0 = T(\partial_{tt}A(0, t)|_{t=1}, z)$ , where

$$\begin{aligned} T(y, z) = & z^3(27z - 4)(64z^2 - 31z + 4)y^3 \\ & - 2z^2(27z - 4)(16z^3 + 39z^2 - 22z + 3)y^2 \\ & + 4(36z^6 + 186z^5 + 118z^4 - 243z^3 + 102z^2 - 17z + 1)y \\ & - 8z^2(z^4 + 8z^3 + 15z^2 - 8z + 1). \end{aligned}$$

The coefficient  $[z^n]\partial_{tt}A(0, t)|_{t=1}$  is the total number of ordered pairs of distinct leaves in the top cells of  $n$ -point dominoes. Since this is more than the total number of leaves and no more than the square of that number, by Proposition 3.10 the dominant singularity of  $\partial_{tt}A(0, t)|_{t=1}$  is at  $4/27$ .

The minimal polynomial  $T(y, z)$  allows us to compute the Puiseux expansion of  $\partial_{tt}A(0, t)|_{t=1}$  at  $z = 4/27$ :

$$\partial_{tt}A(0, t)|_{t=1} = \frac{25}{144} \frac{1}{\sqrt{4/27-z}} + O(1).$$

It follows from [106, Theorem VI.1] that

$$[z^n]\partial_{tt}A(0, t)|_{t=1} = \frac{25}{96} \sqrt{\frac{3}{\pi}} \left(\frac{27}{4}\right)^n n^{-1/2} \left(1 + O\left(\frac{1}{n}\right)\right).$$

Using Stirling's Approximation, we find

$$[z^n]A(0, 1) = |\mathcal{D}_n| = \frac{2(3n+3)!}{(n+2)!(2n+3)!} = \frac{27}{8} \sqrt{\frac{3}{\pi}} \left(\frac{27}{4}\right)^n n^{-5/2} \left(1 + O\left(\frac{1}{n}\right)\right).$$

Thus,

$$\frac{[z^n]\partial_{tt}A(0, t)|_{t=1}}{[z^n]A(0, 1)} = \frac{\frac{25}{96} \sqrt{\frac{3}{\pi}} \left(\frac{27}{4}\right)^n n^{-1/2} \left(1 + O\left(\frac{1}{n}\right)\right)}{\frac{27}{8} \sqrt{\frac{3}{\pi}} \left(\frac{27}{4}\right)^n n^{-5/2} \left(1 + O\left(\frac{1}{n}\right)\right)} = \frac{25}{324} n^2 + O(n).$$

Therefore, the variance is

$$\left(\frac{25}{324} n^2 + O(n)\right) + \left(\frac{5}{18} n + O(1)\right) - \left(\frac{25}{324} n^2 + O(n)\right) = O(n).$$

As the variance is at most linear in  $n$ , the standard deviation is  $O(\sqrt{n})$ . Since the order of the standard deviation is strictly smaller than the order of the expected value, by Chebyshev's inequality the proportion of leaves is concentrated at its mean.  $\square$

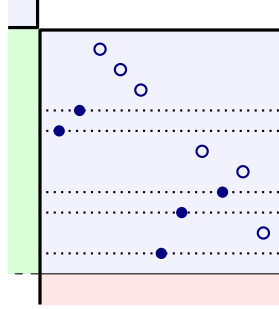


Figure 3.12: Strips in a domino cell.

## Empty strips

In a vertical domino, we consider the cells to be divided into horizontal *strips* by their non-leaf points. For example, in Figure 3.12 the cell is divided into six horizontal strips by its five non-leaf points. We are interested in the number of such strips which contain no leaves, which we call *empty strips*. In Figure 3.12 there are three empty strips.

By the bijection between domino cells and arch systems in Proposition 3.3, empty strips in a 213-avoiding cell correspond to arcs to the left of points that are both the left and right endpoint of an arc (see Figure 3.5). An empty strip is also possible at the bottom of the cell (but not at the top, the uppermost point always being leaf). This possibility does not affect the asymptotics, so we just count *medial* empty strips.

Thus, adapting Proposition 3.5, if  $A(v, s) = A(z, v, s)$  satisfies the functional equation

$$A(v, s) = 1 + zA(v, s) + \frac{z^2 A(v, s)^2}{1 - zsA(v, s)} + z(1+v) \left( A(v, s) + \frac{A(v, s) - A(0, s)}{v} \right),$$

then  $A(0, s) = A(z, 0, s)$  is the bivariate generating function for dominoes in which  $z$  marks points and  $s$  marks medial empty strips in the top cell.

How many empty strips can we expect to find in a domino cell? We calculate the expected number exactly.

**Proposition 3.12.** *The total number of medial empty strips in the top cells of all  $n$ -point dominoes is*

$$\frac{10(3n)!}{(n-3)!(2n+4)!}.$$

*Consequently, the expected number of empty strips in an  $n$ -point domino cell is asymptotically  $5n/27$ .*

*Proof.* The total number of medial empty strips in the top cells of all  $n$ -point dominoes is given by the coefficient of  $z^n$  in  $\partial_s A(0, s)|_{s=1}$ . Using the same approach as in the proof of Proposition 3.10, we can deduce that  $\partial_s A(0, s)|_{s=1}$  is a root of the equation

$$z^4 y^3 - (15z + 2)z^2 y^2 - (10z^3 - 25z^2 + 10z - 1)y - z^3,$$

and verify that the coefficient of  $z^n$  in the power series expansion of  $\partial_s A(0, s)|_{s=1}$  is exactly

$$\frac{10(3n)!}{(n-3)!(2n+4)!}.$$

Therefore, the expected number of medial empty strips in the top cell of a domino with  $n$  points is

$$\frac{\frac{10(3n)!}{(n-3)!(2n+4)!}}{\frac{2(3n+3)!}{(n+2)!(2n+3)!}} = \frac{5n(n-1)(n-2)}{6(3n+1)(3n+2)},$$

from which it follows by symmetry that the expected number of empty strips in an  $n$ -point domino is asymptotically  $5n/27$ .  $\square$

The sequence of coefficients of the power series of  $\partial_s A(0, s)|_{s=1}$  is [A233657](#) in *OEIS* [198]. These are the two-parameter Fuss–Catalan (or Raney) numbers with parameters  $p = 3$  and  $r = 10$ . It would be interesting to find a bijection between medial empty strips in 1324-avoiding dominoes and some other combinatorial class enumerated by this sequence.

Again, we need a concentration result, so we determine the variance.

**Proposition 3.13.** *The proportion of empty strips in the top cell of an  $n$ -point domino is asymptotically concentrated at its mean.*

*Proof.* As before, the minimal polynomial for  $\partial_{ss} A(0, s)|_{s=1}$  is computed from the minimal polynomial for  $\partial_s A(0, s)$ . It is a root of the cubic

$$\begin{aligned} & z^4(27z - 4)(64z^2 - 31z + 4)y^3 \\ & - 2z^2(27z - 4)(64z^4 - 1388z^3 + 534z^2 - 23z - 8)y^2 \\ & - 4(1536z^8 - 22676z^7 + 82275z^6 - 112651z^5 + 72411z^4 - 24430z^3 + 4471z^2 - 421z + 16)y \\ & - 8z^4(64z^5 - 719z^4 + 1371z^3 - 918z^2 + 213z - 16). \end{aligned}$$

This allows us to compute the Puiseux expansion of  $\partial_{ss} A(0, s)|_{s=1}$  at  $z = 4/27$ :

$$\partial_{ss} A(0, s)|_{s=1} = \frac{25}{1296} \frac{1}{\sqrt{4/27-z}} + O(1).$$



It follows that

$$[z^n] \partial_{ss} A(0, s)|_{s=1} = \frac{25}{864} \sqrt{\frac{3}{\pi}} \left(\frac{27}{4}\right)^n n^{-1/2} \left(1 + O\left(\frac{1}{n}\right)\right).$$

Thus,

$$\frac{[z^n] \partial_{ss} A(0, s)|_{s=1}}{[z^n] A(0, 1)} = \frac{\frac{25}{864} \sqrt{\frac{3}{\pi}} \left(\frac{27}{4}\right)^n n^{-1/2} \left(1 + O\left(\frac{1}{n}\right)\right)}{\frac{27}{8} \sqrt{\frac{3}{\pi}} \left(\frac{27}{4}\right)^n n^{-5/2} \left(1 + O\left(\frac{1}{n}\right)\right)} = \frac{25}{2916} n^2 + O(n).$$

Therefore, the variance is

$$\left(\frac{25}{2916} n^2 + O(n)\right) + \left(\frac{5}{54} n + O(1)\right) - \left(\frac{25}{2916} n^2 + O(n)\right) = O(n).$$

The result follows by Chebyshev's inequality.  $\square$

## Dominoes with many leaves and many empty strips

As a consequence of these concentration results, sets of dominoes with many leaves and many empty strips have the same growth rate as the set of all dominoes. For  $\alpha, \beta \in [0, 1]$ , let  $\mathcal{D}_n^{\alpha, \beta}$  be the set of  $n$ -point dominoes with at least  $\alpha n/2$  leaves in each cell and at least  $\beta n/2 + 1$  empty strips in each cell. Let  $\mathcal{D}^{\alpha, \beta} = \bigcup_n \mathcal{D}_n^{\alpha, \beta}$ . The use of  $\beta n/2 + 1$ , rather than  $\beta n/2$ , is explained in the proof of Proposition 3.15 below.

**Corollary 3.14.** *If  $\alpha < 5/9$  and  $\beta < 5/27$ , then  $\text{gr}(\mathcal{D}^{\alpha, \beta}) = 27/4$ .*

*Proof.* By Propositions 3.10 and 3.11, for sufficiently large  $n$ , at least four fifths of  $n$ -point dominoes have  $\alpha n/2$  or more leaves in their top cell, and, by symmetry, at least four fifths have  $\alpha n/2$  or more leaves in their bottom cell. Let  $\beta'$  be in the open interval  $(\beta, 5/27)$ . Then, for sufficiently large  $n$ , we have  $\beta' n/2 \geq \beta n/2 + 1$  and so, applying Propositions 3.12 and 3.13 with  $\beta'$ , at least four fifths of  $n$ -point dominoes have  $\beta n/2 + 1$  or more empty strips in their top cell, and at least four fifths have  $\beta n/2 + 1$  or more empty strips in their bottom cell. Hence asymptotically, at least one fifth of all dominoes are in  $\mathcal{D}^{\alpha, \beta}$ . The result follows from Theorem 3.2.  $\square$

An analogous result holds for sets of *balanced* dominoes with many leaves and many empty strips. Let  $\mathcal{B}_m^{\alpha, \beta}$  be the set of  $2m$ -point balanced dominoes, with at least  $\alpha m$  leaves in each cell and at least  $\beta m + 1$  empty strips in each cell, and let  $\mathcal{B}^{\alpha, \beta} = \bigcup_m \mathcal{B}_m^{\alpha, \beta}$ .

**Proposition 3.15.** *If  $\alpha < 5/9$  and  $\beta < 5/27$ , then  $\text{gr}(\mathcal{B}^{\alpha, \beta}) = 27/4$ .*

The proof mirrors that of Proposition 3.7.

*Proof.* For suitable values of the parameters, let  $\mathcal{L}(t, b, \ell_{\top}, \ell_{\text{B}}, e_{\top}, e_{\text{B}})$  denote the set of  $(t + b)$ -point dominoes with  $t$  points in the top cell,  $b$  points in the bottom cell,  $\ell_{\top}$  leaves in the top cell,  $\ell_{\text{B}}$  leaves in the bottom cell,  $e_{\top}$  empty strips in the top cell and  $e_{\text{B}}$  empty strips in the bottom cell. For a given  $m$ , let  $\mathcal{L}_m$  be some such set whose size is maximal subject to the conditions  $t + b = m$ ,  $\ell_{\top}, \ell_{\text{B}} \geq \alpha m/2$  and  $e_{\top}, e_{\text{B}} \geq \beta m/2 + 1$ . Note that  $\mathcal{L}_m \subseteq \mathcal{D}_m^{\alpha, \beta}$ . Since  $0 \leq t, \ell_{\top}, \ell_{\text{B}}, e_{\top}, e_{\text{B}} \leq m$ , there are at most  $(m + 1)^5$  possible choices for the parameters. Hence by the pigeonhole principle,

$$|\mathcal{L}_m| \geq \frac{|\mathcal{D}_m^{\alpha, \beta}|}{(m + 1)^5}.$$

Let  $\sigma$  and  $\tau$  be any two  $m$ -point dominoes from  $\mathcal{L}_m$ . Consider the domino  $\rho = \sigma \otimes \hat{\tau}$ , whose arch configuration is constructed by concatenating the arch configuration of  $\sigma$  and the arch configuration of the  $180^\circ$  rotation of  $\tau$ . This domino has  $m$  points in each cell and  $\ell_{\top} + \ell_{\text{B}} \geq \alpha m$  leaves in each cell. If the top cell of  $\sigma$  has an empty strip at the bottom, then this combines with the non-empty strip at the bottom of the bottom cell of  $\tau$ , in which case the top cell of  $\rho$  has  $e_{\top} + e_{\text{B}} - 1$  empty strips. Otherwise it has  $e_{\top} + e_{\text{B}}$  empty strips. In either case, this is at least  $\beta m + 1$ . An analogous argument applies to the bottom cell, so  $\rho$  is a balanced domino in  $\mathcal{B}_m^{\alpha, \beta}$ .

Moreover,  $\sigma$  and  $\tau$  can be recovered from  $\rho$  simply by splitting its arch configuration into two halves. Thus,

$$|\mathcal{B}_m^{\alpha, \beta}| \geq |\mathcal{L}_m|^2 \geq \frac{|\mathcal{D}_m^{\alpha, \beta}|^2}{(m + 1)^{10}}.$$

Since it is also the case that  $|\mathcal{D}_{2m}^{\alpha, \beta}| \geq |\mathcal{B}_m^{\alpha, \beta}|$ , it follows, by taking the  $2m$ th root, and the limit as  $m$  tends to infinity, that  $\text{gr}(\mathcal{B}^{\alpha, \beta}) = \text{gr}(\mathcal{D}^{\alpha, \beta}) = 27/4$ .  $\square$

## A better lower bound

In this final section, we modify the construction used to prove Theorem 3.9 to yield an improved lower bound. We make use of exactly the same decomposition of the staircase, which we reproduce here in Figure 3.13. However, we change the rules concerning the permitted interleaving of points between the cells. We also exploit the additional properties of dominoes established in Section 3.7.

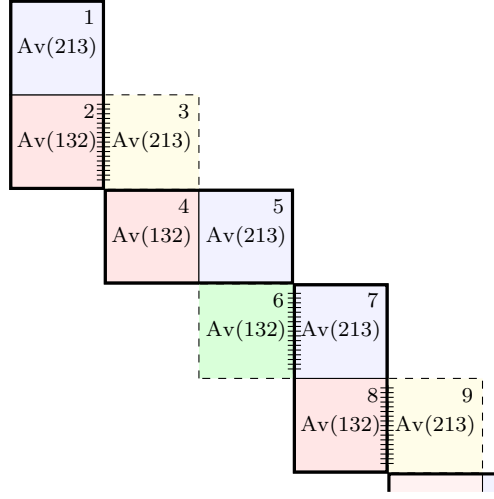


Figure 3.13: The decomposition of the staircase into dominoes and connecting cells.

Recall that in our earlier construction, we ensure that there is no occurrence of 1324 by requiring every point in a domino cell to be positioned between the components in the adjacent connecting cells, as illustrated in Figure 3.11. For our improved lower bound, we relax this restriction in the case of domino cells *to the left or right of a connecting cell*. In this case, we require only that *non-leaves* in a domino cell are positioned between the components. Leaves may be positioned arbitrarily. See Figure 3.14 for an illustration of a 132-avoiding connecting cell and its adjacent domino cells. In the domino cell to the right, leaves are shown as rings and non-leaves as disks.

This still prevents any occurrence of 1324. For example, if a domino cell is to the right of a connecting cell, then this restriction ensures that in any occurrence of 132 with the 13 in the connecting cell and the 2 in the domino cell the 2 is a leaf, so there can be no point to its upper right to complete a 1324. In Figure 3.13, this greater freedom is shown using small lines between connecting cells and horizontally adjacent domino cells. Observe that this flexibility *only* applies to the cells of vertical dominoes in the decomposition. We could similarly relax the restriction in the case of domino cells above and below a connecting cell. However, this results in a structure we have been unable to analyse.

This refined construction enables us to establish an improved lower bound on the growth rate of 1324-avoiders.

**Theorem 3.16.** *The growth rate of  $\text{Av}(1324)$  is at least 10.271012.*

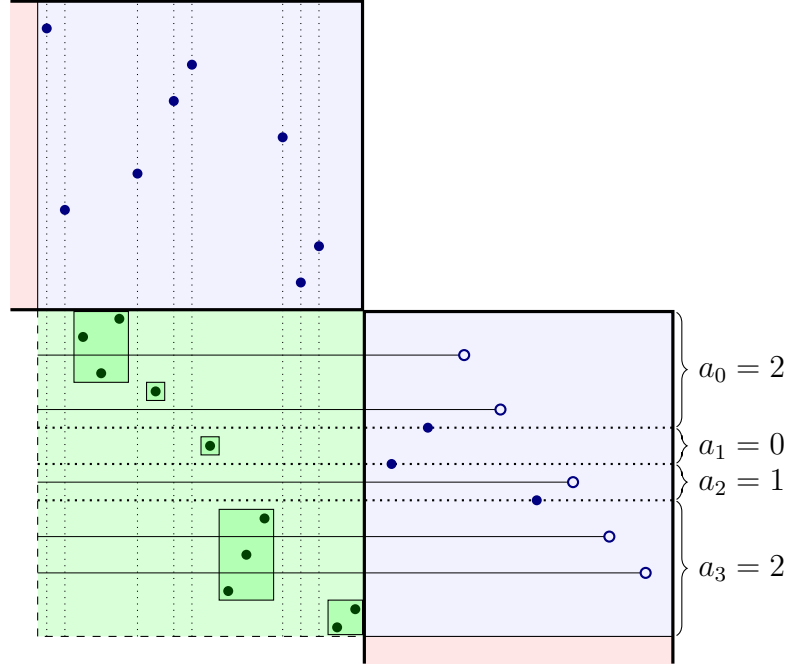


Figure 3.14: Interleaving the points in a connecting cell with those in two domino cells.

## Horizontally interleaved connecting cells

Let us consider how a connecting cell can be interleaved with a horizontally adjacent domino cell. We want to enumerate diagrams like the lower two cells of Figure 3.14, where the points in the domino cell at the right have been erased, but the horizontal lines, solid for leaves and dotted for non-leaves, have been retained to record the positions of the points relative to the points in the connecting cell. Let us call these configurations *horizontally interleaved connecting cells*.

We begin with the generating function for connecting cells,

$$H(z, q) = \frac{1}{1 - qQ(z)} = \frac{2}{2 - q + q\sqrt{1 - 4z}}, \quad (3.3)$$

where  $z$  marks points,  $q$  marks components, and  $Q(z) = \frac{1}{2}(1 - \sqrt{1 - 4z})$  is the generating function for components of a connecting cell.

As described in Section 3.7.2, the non-leaves of a vertical domino cell divide it and the adjacent connecting cell into horizontal strips. Suppose that such a domino cell has  $\ell$  leaves and  $r$  non-leaves. The  $r$  non-leaves divide the cell into  $r + 1$  horizontal strips, each containing a certain number

of leaves. Let  $a_i$  denote the number of leaves in the  $i$ th strip from the top, for  $i = 0, \dots, r$ , so  $a_0 + \dots + a_r = \ell$ . See Figure 3.14 for an illustration.

The generating function for the possibilities in the  $i$ th strip is given by

$$H_{a_i}(z, q) = \Omega_{a_i}[H(z, q)], \quad (3.4)$$

where each  $\Omega_j$  is a linear operator given by

$$\Omega_j[z^n] = \binom{n+j}{j} z^n,$$

or equivalently,

$$\Omega_j[F(z)] = \frac{1}{j!} \frac{\partial^j}{\partial z^j} (z^j F(z)). \quad (3.5)$$

Hence, for a fixed sequence  $(a_i)_{i=0}^r$  of strip sizes, the generating function for horizontally interleaved connecting cells, counting once each possible way of interleaving with the contents of the horizontally adjacent domino cell, is given by

$$\prod_{i=0}^r H_{a_i}(z, q). \quad (3.6)$$

We cannot work directly with this expression, since it would require us to keep track of all the strip sizes. So, in order to establish a lower bound, we seek to minimise the above expression over all sequences  $a_0, \dots, a_r$  such that  $a_0 + \dots + a_r = \ell$ . With the next two propositions we demonstrate that such a minimum exists for any fixed  $r$  and  $\ell$ , in the sense that every coefficient of (3.6) is minimised for the same sequence  $a_0, \dots, a_r$ . More specifically, we prove that this minimum occurs when no two terms of the sequence differ by more than 1. We call such a sequence *equitable*.

In our refinement of the staircase, a certain number of the strips are required to be empty. With this additional requirement, for a lower bound, we thus need an equitable distribution of the leaves among the rest of the strips.

The following proposition is framed in the general setting of partially ordered rings, though for our purposes these are always rings of formal power series with real coefficients. Recall that a *partially ordered ring*  $(R, \leq)$ , is a (commutative) ring  $R$  together with a partial order  $\leq$  on the elements of  $R$  such that if  $a, b, c \in R$  then  $a \leq b$  if and only if  $a + c \leq b + c$ , and  $a, b \geq 0$  implies  $ab \geq 0$ . Given such a ring  $(R, \leq)$ , we define  $(R[[q]], \leq)$  to be the ring of formal power series over  $R$  equipped with the partial order defined by  $h(q) \geq 0$  if and only if every coefficient of  $h(q)$  is in  $R_{\geq 0} = \{r \in R : r \geq 0\}$ .

A sequence  $a_0, a_1, \dots$  in  $(R, \leq)$  is *log-convex* if, for every pair of integers  $i, j$  with  $0 \leq i < j$ , we have  $a_i a_{j+1} \geq a_{i+1} a_j$ .

**Proposition 3.17.** *Let  $(R, \leq)$  be a partially ordered ring and let  $a_0, a_1, \dots$  be a log-convex sequence in  $R_{\geq 0}$ . Furthermore, let  $F(z) = a_0 + a_1 z + \dots$  be the generating function of this sequence. Then the sequence  $\Omega_0[F(z)], \Omega_1[F(z)], \dots$  is log-convex in the partially ordered ring  $(R[[z]], \leq)$ .*

*Proof.* We just need to show that for each  $k \geq 0$  and each  $a > b \geq 0$ ,

$$[z^k] \left( \Omega_{a+1}[F(z)] \Omega_b[F(z)] - \Omega_a[F(z)] \Omega_{b+1}[F(z)] \right) \geq 0.$$

This coefficient can be computed as

$$\begin{aligned} & \sum_{j=0}^k \binom{a+1+j}{j} a_j \binom{b+k-j}{k-j} a_{k-j} - \sum_{j=0}^k \binom{a+j}{j} a_j \binom{b+1+k-j}{k-j} a_{k-j} \\ &= \sum_{j=0}^k \left( \binom{a+1+j}{j} \binom{b+k-j}{k-j} - \binom{b+1+j}{j} \binom{a+k-j}{k-j} \right) a_j a_{k-j} \\ &= \sum_{j=0}^k \sum_{i=0}^j \left( \binom{a+i}{i} \binom{b+k-j}{k-j} - \binom{b+i}{i} \binom{a+k-j}{k-j} \right) a_j a_{k-j} \\ &= \sum_{j=0}^k \sum_{i=0}^{k-j} \left( \binom{a+i}{i} \binom{b+j}{j} - \binom{b+i}{i} \binom{a+j}{j} \right) a_j a_{k-j} \\ &= \sum_{j=0}^k \sum_{i=0}^{\min(j-1, k-j)} \left( \binom{a+i}{i} \binom{b+j}{j} - \binom{b+i}{i} \binom{a+j}{j} \right) (a_j a_{k-j} - a_i a_{k-i}). \end{aligned}$$

Now, the coefficient of  $a_j a_{k-j} - a_i a_{k-i}$  in each summand, namely

$$\binom{a+i}{i} \binom{b+j}{j} - \binom{b+i}{i} \binom{a+j}{j},$$

is negative, since  $i < j$  and  $a > b$ . Also, since  $i \leq k-j$ , we have  $a_i a_{k-i} \geq a_j a_{k-j}$ . Hence each summand is nonnegative and the entire sum is positive, which implies that the sequence  $\Omega_0[F(z)], \Omega_1[F(z)], \dots$  is log-convex in  $(R[[z]], \leq)$ .  $\square$

We now apply this to the enumeration of horizontally interleaved connecting cells.

**Proposition 3.18.** *Let*

$$H(z, q) = \frac{2}{2 - q + q\sqrt{1 - 4z}} = h_0(q) + zh_1(q) + z^2 h_2(q) + \dots$$

be the generating function for connecting cells where  $z$  marks points and  $q$  marks components. Then the sequence of polynomials  $h_0(q), h_1(q), \dots$  is log-convex in  $(\mathbb{R}[[q]], \leq)$ . Consequently, the sequence  $H(z, q), H_1(z, q), H_2(z, q), \dots$  is log-convex in  $(\mathbb{R}[[z, q]], \leq)$ .

*Proof.* Since the generating function  $H(z, q)$  satisfies the equation

$$H(z, q) = 1 + z \frac{q^2 H(z, q) - q H(z, 1)}{q - 1},$$

it follows that for each  $i \geq 1$ ,

$$h_i(q) = \frac{q^2 h_{i-1}(q) - q h_{i-1}(1)}{q - 1} = \frac{q^2 h_{i-1}(q) - q c_{i-1}}{q - 1},$$

where  $c_n = \binom{2n}{n} / (n + 1)$  is the  $n$ th Catalan number. Rearranging this gives the equation

$$h_{i-1}(q) = \frac{(q - 1)h_i(q) + q c_{i-1}}{q^2}.$$

We need to prove that if  $j > i \geq 1$  then we have  $h_{i-1}(q)h_j(q) \geq h_i(q)h_{j-1}(q)$ . This happens if and only if

$$\frac{(q - 1)h_i(q) + q c_{i-1}}{q^2} h_j(q) \geq \frac{(q - 1)h_j(q) + q c_{j-1}}{q^2} h_i(q),$$

which simplifies to

$$c_{i-1}h_j(q) - c_{j-1}h_i(q) \geq 0.$$

One can easily prove by induction, or otherwise, that

$$h_i(q) = \sum_{k=1}^i h_{i,k} q^k, \quad \text{where } h_{i,k} = \frac{k}{2i - k} \binom{2i - k}{i}.$$

It suffices to demonstrate that  $c_{i-1}h_{j,k} - c_{j-1}h_{i,k} \geq 0$  whenever  $j > i \geq k \geq 1$ . By transitivity, we only need consider the case  $j = i + 1$ , when it is readily confirmed that the required inequality holds:

$$c_{i-1}h_{i+1,k} - c_i h_{i,k} = \frac{k(k-1)(k-2)(2i-2)!(2i-k-1)!}{(i+1)! i! (i-1)! (i-k+1)!} \geq 0, \quad \text{if } i \geq k \geq 1.$$

Hence, the sequence  $h_0(q), h_1(q), \dots$  is log-convex in  $(\mathbb{R}[[q]], \leq)$ . Consequently, by Proposition 3.17, the sequence  $H(z, q), H_1(z, q), H_2(z, q), \dots$  is log-convex in  $(\mathbb{R}[[z, q]], \leq)$ .  $\square$

Thus, as claimed above, among all sequences  $a_0, \dots, a_r$  which satisfy  $a_0 + \dots + a_r = \ell$ , the minimum value of every coefficient of

$$\prod_{i=0}^r H_{a_i}(z, q)$$

is achieved by *equitable* sequences, that is in which  $|a_i - a_j| \leq 1$  for every  $i, j \in \{0, \dots, r\}$ . This, therefore, is what we apply to the non-empty strips to give a lower bound for the number of horizontally interleaved connecting cells.

## Refining the staircase

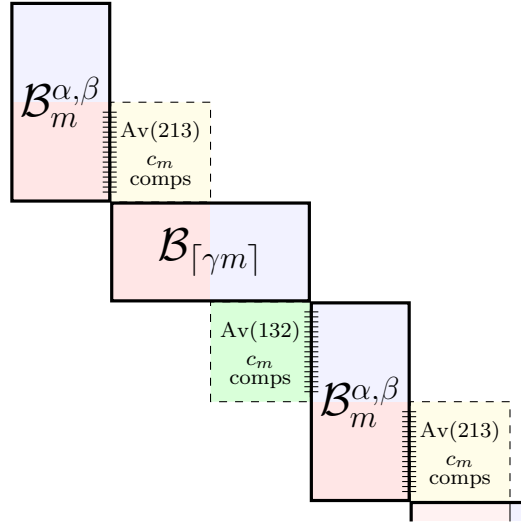


Figure 3.15: The scheme used to calculate the improved lower bound.

We are now ready to describe more precisely how we modify our construction so as to yield an improved lower bound. This description is accompanied by Figure 3.15. Recall that  $\mathcal{B}^{\alpha, \beta} = \bigcup_m \mathcal{B}_m^{\alpha, \beta}$ , where  $\mathcal{B}_m^{\alpha, \beta}$  consists of dominoes in which each cell has  $m$  points, at least  $\alpha m$  leaves and at least  $\beta m + 1$  empty strips. Let  $\alpha, \beta > 0$  be sufficiently small that  $\mathcal{B}^{\alpha, \beta}$  has exponential growth rate  $27/4$ . By Proposition 3.15, we may choose any  $\alpha < 5/9$  and  $\beta < 5/27$ . We also require that  $\alpha \geq 11/20$  and  $\beta \geq 7/40$ .

For fixed values of parameters  $\alpha, \beta, \gamma$  and  $\kappa$ , and sufficiently large  $k$  and  $m$ , let  $\mathcal{P}_{k, m}$  be the set of gridded permutations, gridded in the first  $6k + 2$  cells of the staircase, satisfying the following conditions.



- Each non-leaf in a cell of a vertical domino is positioned between components of the horizontally adjacent connecting cell.
- Each point in a cell of a horizontal domino is positioned between components of the vertically adjacent connecting cell.
- Each vertical domino is an element of  $\mathcal{B}_m^{\alpha,\beta}$ .
- Each horizontal domino is a balanced domino with  $\lceil \gamma m \rceil$  points in each cell, for some  $\gamma > 0$  to be chosen later.
- Each connecting cell has  $c_m$  components, where  $\lim_{m \rightarrow \infty} c_m/m = \kappa$ , for some  $\kappa > 0$ ; the value of  $\kappa$  and the sequence  $(c_m)$  are to be chosen later.

Note that each domino cell contains a fixed number of points (either  $m$  or  $\lceil \gamma m \rceil$ ). However, the number of points in a connecting cell is not fixed, although its number of skew indecomposable components,  $c_m$ , is.

We begin by establishing a lower bound for the enumeration of horizontally interleaved connecting cells in  $\mathcal{P}_{k,m}$ . At least  $\lceil \alpha m \rceil$  of the points are leaves, and at least  $\lceil \beta m \rceil + 1$  of the strips are empty. Note first that changing a non-leaf to a leaf can only increase the number of ways of performing the interleaving. So, for a lower bound, we may assume there are exactly  $\lceil \alpha m \rceil$  leaves. Note also that, since  $\alpha > 1/2$ , an equitable distribution of leaves among the strips allocates at least one leaf to each strip. Hence, any increase in the number of empty strips can only make the distribution less equitable. So, for a lower bound, we may assume there are exactly  $\lceil \beta m \rceil + 1$  empty strips.

With these assumptions, given  $\alpha$  in the interval  $[11/20, 5/9]$ ,  $\beta$  in the interval  $[7/40, 5/27]$  and  $m \geq 32$ , an equitable distribution of the leaves among the non-empty strips consists of

- $e_0(m) = \lceil \beta m \rceil + 1$  empty strips,
- $e_2(m) = 3m - 4\lceil \alpha m \rceil - 3\lceil \beta m \rceil$  two-leaf strips, and
- $e_3(m) = 3\lceil \alpha m \rceil + 2\lceil \beta m \rceil - 2m$  three-leaf strips.

The expressions for  $e_2(m)$  and  $e_3(m)$  are the solutions of the equations

$$\begin{aligned} 2e_2(m) + 3e_3(m) &= \lceil \alpha m \rceil, \\ e_0(m) + e_2(m) + e_3(m) &= m - \lceil \alpha m \rceil + 1, \end{aligned}$$

for the total number of leaves and the total number of strips, respectively. The bounds on  $\alpha$ ,  $\beta$  and  $m$  ensure that each of the  $e_j(m)$  is nonnegative.

Thus, since the number of components in each connecting cell is exactly  $c_m$ ,

$$J_m(z) = [q^{c_m}] (H(z, q)^{e_0(m)} H_2(z, q)^{e_2(m)} H_3(z, q)^{e_3(m)}) \quad (3.7)$$

is a lower bound for the generating function of horizontally interleaved connecting cells in  $\mathcal{P}_{k,m}$ .

To understand the asymptotics of  $J_m(z)$  for large  $m$ , we use the following general result, concerning the exponential growth rate of combinatorial objects whose generating function has coefficients of the form  $[x^{(\kappa+o(1))n}] \prod_{j=1}^r F_j(x)^{(\alpha_j+o(1))n}$ , for some fixed  $\alpha_1, \dots, \alpha_r$  and  $\kappa$ .

**Lemma 3.19.** *Let  $\alpha_1, \dots, \alpha_r$  and  $\kappa$  be positive constants. For each  $j \in [r]$ , let  $F_j(x)$  be a power series with radius of convergence  $\rho_j$ . For each  $j$ , suppose that  $a_{j,1}, a_{j,2}, \dots$  is a sequence of positive integers such that  $\lim_{n \rightarrow \infty} a_{j,n}/n = \alpha_j$ , and that there is some positive  $x_0$ , smaller than every  $\rho_j$ , satisfying*

$$x_0 \sum_{j=1}^r \alpha_j \frac{F'_j(x_0)}{F_j(x_0)} = \kappa.$$

*Then there exists a sequence of positive integers  $c_1, c_2, \dots$  such that  $\lim_{n \rightarrow \infty} c_n/n = \kappa$ , for which*

$$\lim_{n \rightarrow \infty} \left( [x^{c_n}] \prod_{j=1}^r F_j(x)^{a_{j,n}} \right)^{1/n} = x_0^{-\kappa} \prod_{j=1}^r F_j(x_0)^{\alpha_j}.$$

This lemma is rather easier to understand and its proof easier to follow when  $r = 1$ . Unfortunately, we need the more general version.

*Proof.* For each  $j$ , define the probability generating function

$$G_j(x) = \frac{F_j(x_0 x)}{F_j(x_0)}.$$

This definition is valid because  $x_0 < \rho_j$ .

The corresponding expected value is  $\mu_j = G'_j(1) = x_0 F'_j(x_0)/F_j(x_0)$ , so

$$\sum_{j=1}^r \alpha_j \mu_j = \kappa.$$

For each  $j$ , let  $X_j$  be a random variable with probability generating function  $G_j$ . For each  $n > 0$ , let  $Y_n$  be the random variable defined by adding  $a_{j,n}$  independent samples from  $X_j$  for each  $j$ . Then the expected value  $\lambda_n$  of  $Y_n$  is given by

$$\lambda_n = \sum_{j=1}^r a_{j,n} \mu_j.$$

Moreover, it follows from the law of large numbers that if  $\varepsilon > 0$ , then the probability  $p_{\varepsilon,n}$  that  $Y_n$  lies in the interval  $(\lambda_n(1-\varepsilon), \lambda_n(1+\varepsilon))$  converges to 1 as  $n$  tends to infinity. In terms of generating functions, this means

$$\lim_{n \rightarrow \infty} \sum_{c \in (\lambda_n(1-\varepsilon), \lambda_n(1+\varepsilon))} [x^c] \prod_{j=1}^r G_j(x)^{a_{j,n}} = \lim_{n \rightarrow \infty} p_{\varepsilon,n} = 1. \quad (3.8)$$

For each pair  $\varepsilon, n$ , let  $c(\varepsilon, n)$  be the value in the interval  $(\lambda_n(1-\varepsilon), \lambda_n(1+\varepsilon))$  which maximises

$$[x^{c(\varepsilon,n)}] \prod_{j=1}^r G_j(x)^{a_{j,n}}.$$

Then, by (3.8), we have

$$\liminf_{n \rightarrow \infty} 2\varepsilon \lambda_n [x^{c(\varepsilon,n)}] \prod_{j=1}^r G_j(x)^{a_{j,n}} \geq 1.$$

It follows that

$$\lim_{n \rightarrow \infty} \left( [x^{c(\varepsilon,n)}] \prod_{j=1}^r G_j(x)^{a_{j,n}} \right)^{1/n} = 1.$$

Therefore, we can choose a sequence  $c_1, c_2, \dots$  by setting  $c_n = c(\varepsilon_n, n)$  in such a way that

$$\lim_{n \rightarrow \infty} \varepsilon_n = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \left( [x^{c_n}] \prod_{j=1}^r G_j(x)^{a_{j,n}} \right)^{1/n} = 1.$$

We now show that this sequence satisfies the desired properties. First note that  $c_n$  lies in the interval  $(\lambda_n(1-\varepsilon_n), \lambda_n(1+\varepsilon_n))$ , so the ratio  $c_n/\lambda_n$  converges to 1. Moreover,

$$\lim_{n \rightarrow \infty} \lambda_n/n = \lim_{n \rightarrow \infty} \sum_{j=1}^r a_{j,n} \mu_j/n = \sum_{j=1}^r \alpha_j \mu_j = \kappa.$$

Hence, the ratio  $c_n/n$  converges to  $\kappa$ . Finally,

$$\begin{aligned} \lim_{n \rightarrow \infty} \left( [x^{c_n}] \prod_{j=1}^r F_j(x)^{a_{j,n}} \right)^{1/n} &= \lim_{n \rightarrow \infty} \left( [x^{c_n}] \prod_{j=1}^r G_j(x/x_0)^{a_{j,n}} \prod_{j=1}^r F_j(x_0)^{a_{j,n}} \right)^{1/n} \\ &= x_0^{-\kappa} \lim_{n \rightarrow \infty} \left( [x^{c_n}] \prod_{j=1}^r G_j(x)^{a_{j,n}} \right)^{1/n} \prod_{j=1}^r F_j(x_0)^{\alpha_j} \\ &= x_0^{-\kappa} \prod_{j=1}^r F_j(x_0)^{\alpha_j}. \end{aligned} \quad \square$$

Let us apply this lemma to  $J_m(z)$ , as defined in (3.7). For any fixed  $z_0$ , there exists a sequence of positive integers  $c_1, c_2, \dots$  such that  $\lim_{m \rightarrow \infty} c_m/m = \kappa$ , for which

$$\lim_{m \rightarrow \infty} J_m(z_0)^{1/m} = q_0^{-\kappa} H(z_0, q_0)^\beta H_2(z_0, q_0)^{3-4\alpha-3\beta} H_3(z_0, q_0)^{3\alpha+2\beta-2}, \quad (3.9)$$

where  $q_0 = q_0(z_0)$  satisfies

$$\beta \frac{\frac{d}{dq} H(z_0, q)}{H(z_0, q)} \Big|_{q=q_0} + (3-4\alpha-3\beta) \frac{\frac{d}{dq} H_2(z_0, q)}{H_2(z_0, q)} \Big|_{q=q_0} + (3\alpha+2\beta-2) \frac{\frac{d}{dq} H_3(z_0, q)}{H_3(z_0, q)} \Big|_{q=q_0} = \frac{\kappa}{q_0}, \quad (3.10)$$

as long as  $q_0$  is less than the radius of convergence in  $q$  of the  $H_j(z_0, q)$ . Note that each  $H_j(z, q)$  can be determined explicitly from the definitions in (3.3), (3.4) and (3.5).

## Enumerating the refined staircase

The first  $6k + 2$  cells of the staircase consist of a total of  $k + 1$  vertical dominoes, each in  $\mathcal{B}_m^{\alpha, \beta}$ , a total of  $k$  horizontal dominoes, each in  $\mathcal{B}_{\lceil \gamma m \rceil}$ , and  $2k$  connecting cells. Thus, for sufficiently large  $m$ , the generating function for  $\mathcal{P}_{k,m}$  is bounded below by

$$F_{k,m}(z) = \left| \mathcal{B}_m^{\alpha, \beta} \right|^{k+1} z^{2m(k+1)} \left| \mathcal{B}_{\lceil \gamma m \rceil} \right|^k z^{2\lceil \gamma m \rceil k} J_m(z)^{2k} \binom{\lceil \gamma m \rceil + c_m}{c_m}^{2k},$$

where the final binomial coefficient counts the number of possible ways of interleaving the  $\lceil \gamma m \rceil$  points in a horizontal domino cell with the  $c_m$  components in a vertically adjacent connecting cell.

Let  $A(z)$  be the generating function for  $\text{Av}(1324)$ , and for each  $k$ , let  $A_k(z)$  be the generating function for the set of 1324-avoiding gridded permutations in the first  $6k + 2$  cells of the (original) staircase. Thus, for any fixed  $k$  and  $m$ , and all  $n$ ,

$$[z^n] F_{k,m}(z) \leq [z^n] A_k(z) \leq \binom{n + 6k + 1}{6k + 1} [z^n] A(z).$$

So, since the binomial coefficient is a polynomial in  $n$ , it follows from the second inequality that the radius of convergence of  $A_k(z)$  is at least that of  $A(z)$ .

Hence, for any  $k$ , and any fixed  $z_0$  within the radius of convergence of  $A(z)$ , the value of  $F_{k,m}(z_0)$  is bounded above by  $A_k(z_0)$  for every  $m$ . So

$\limsup_{m \rightarrow \infty} F_{k,m}(z_0)^{1/m} \leq 1$ , and as a consequence,

$$\lim_{k \rightarrow \infty} \left( \limsup_{m \rightarrow \infty} F_{k,m}(z_0)^{1/m} \right)^{1/2k} \leq 1.$$

By Propositions 3.7 and 3.15, equation (3.9) and Stirling's approximation, the left side of this inequality is equal to

$$G(z_0) = \left( \frac{27z_0}{4} \right)^{1+\gamma} q_0^{-\kappa} H(z_0, q_0)^\beta H_2(z_0, q_0)^{3-4\alpha-3\beta} H_3(z_0, q_0)^{3\alpha+2\beta-2} \frac{(\gamma + \kappa)^{\gamma+\kappa}}{\gamma^\gamma \kappa^\kappa},$$

for some appropriate sequence  $c_1, c_2, \dots$ , where  $q_0$  is defined by (3.10).

To prove Theorem 3.16, it now suffices to find suitable values of  $\alpha, \beta, \gamma, \kappa$  and  $z_0$ , for which  $G(z_0) > 1$  and such that  $q_0$  satisfying (3.10) is less than the radius of convergence in  $q$  of the  $H_j(z_0, q)$ . Any such  $z_0$  lies outside the radius of convergence of  $A(z)$  and so  $1/z_0$  is a lower bound on the growth rate of  $\text{Av}(1324)$ . We thus seek  $z_0$  as small as possible.

Using  $\alpha = 5/9 - 10^{-8}$ ,  $\beta = 5/27 - 10^{-8}$ ,  $\gamma \approx 0.951509$  and  $\kappa \approx 0.496339$ , we may take the value of  $z_0$  to be approximately 0.097361383. Then  $q_0 \approx 2.917054$  and the radius of convergence of the  $H_j(z_0, q)$  is about 9.15, so  $q_0$  is in the required range, and  $G(z_0) > 1$ . Therefore  $1/z_0 \approx 10.271012$  is a lower bound on the growth rate of  $\text{Av}(1324)$ .<sup>3</sup>

## Improving the lower bound further

How might this result be improved? Firstly, if we determined the expected proportion of  $k$ -leaf strips for  $k \geq 1$ , and established that their distribution was concentrated, then that would affect the optimal distribution of points between the strips, leading to a better bound. It is possible to modify the functional equation for dominoes to record  $k$ -leaf strips, for any  $k$ , but the result is complicated and it has not been possible to analyse the result, even for  $k = 1$ .

Secondly, as mentioned at the beginning of Section 3.8, we could relax our construction to permit leaves in vertically adjacent domino cells to be positioned arbitrarily, like the leaves in horizontally adjacent domino cells are. Due to the complex interaction between the interleaving of points in two directions, we have not been able to determine a lower bound for the number of possibilities. It seems likely that the one-dimensional solution in

---

<sup>3</sup>An alternative approach to analysing the refined staircase suggests that we could take the lower bound to be an algebraic number with a minimal polynomial of degree 104, whose value is approximately 10.27101292824530.

which leaves are distributed equitably between the strips does not carry over to interleaving in two directions.

Finally, if we established (a lower bound on) the growth rate of the set of permutations gridded in the first three cells of the staircase, then we could decompose the staircase into three-celled *trominoes* to yield a new bound. However, enumerating trominoes seems to require some new ideas.

# Chapter 4

## Enumerating planar Eulerian orientations

### Introduction

A *planar map* is a connected planar graph embedded in the sphere, and taken up to orientation preserving homeomorphism (see Figure 4.1). The enumeration of planar maps is a venerable topic in combinatorics, which was born in the early sixties with the pioneering work of William Tutte [200, 201]. Fifteen years later it started a second, independent, life in theoretical physics, where planar maps are seen as a discrete model of *quantum gravity* [55, 26]. The enumeration of maps also has connections with factorizations of permutations, and hence representations of the symmetric group [131, 132]. Finally, 40 years after the first enumerative results of Tutte, planar maps crossed the border between combinatorics and probability theory, where they are now studied as random metric spaces [8, 64, 146, 153]. The limit behaviour of large planar random maps is now well understood, and gave birth to a variety of limiting objects, either continuous like the Brownian map [72, 146, 147, 157], or discrete like the UIPQ (uniform infinite planar quadrangulation) [8, 63, 73, 155].

The enumeration of maps equipped with some additional structure (a spanning tree, a proper colouring, a self-avoiding-walk, a configuration of the Ising model...) has attracted the interest of both combinatorialists and theoretical physicists since the early days of this study [80, 136, 159, 202, 203]. At the moment, a challenge in both physics and probability theory is to understand the limiting behaviour of maps equipped with one such structure [44, 135, 138, 175, 182].

The enumeration of these “decorated” maps, and understanding their structure, remain the very first building blocks towards the resolution of such

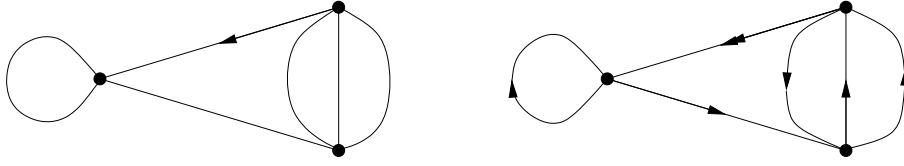


Figure 4.1: Left: a rooted planar map, which is 4-valent (or: quartic). Right: the same map, equipped with an Eulerian orientation.

challenges. Recently, the natural question of counting maps equipped with an *Eulerian orientation* (where the edges are oriented in such a way that every vertex has as many incoming as outgoing edges, see Figure 4.1) was raised by Bonichon, Bousquet-Mélou, Dorbec and Pennarun [42]. They did not solve the problem, but gave sequences of lower bounds and upper bounds on the number of planar Eulerian orientations. They also suggested studying the special case of 4-valent (or *quartic*) Eulerian orientations. This problem had already been studied around 2000 in theoretical physics, where it coincides with the *ice model on a random lattice* [143, 214]. Another fact that makes this case particularly relevant is that the number of such orientations with  $n$  vertices is known to be the number of 3-coloured quadrangulations with  $n$  faces [211].

In this chapter we start by presenting numerical work which allowed us to conjecture the asymptotic form of the two generating functions  $Q(t)$ , which counts quartic planar Eulerian orientations by vertices and  $G(t)$ , which counts planar Eulerian orientations by edges. These lead us to more specific conjectures about the exact, D-algebraic form of both generating functions. We then prove these conjectures, thus completely solving these two enumeration problems.

Let us now state our main two theorems. As is usual with maps, our orientations are *rooted*, which means that we mark one (oriented) edge (Figure 4.1, right). Orientations of small size are shown in Figure 4.2.

**Theorem 4.1.** *Let  $R(t) \equiv R$  be the unique formal power series with constant term 0 satisfying*

$$t = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n} \binom{3n}{n} R^{n+1}.$$

*Then the generating function of quartic rooted planar Eulerian orientations, counted by vertices, is*

$$Q(t) = \frac{1}{3t^2} (t - 3t^2 - R(t)).$$

*This is a differentially algebraic series, satisfying a non-linear differential equation of order 2 whose coefficients are polynomials in  $t$ . The number  $q_n$*



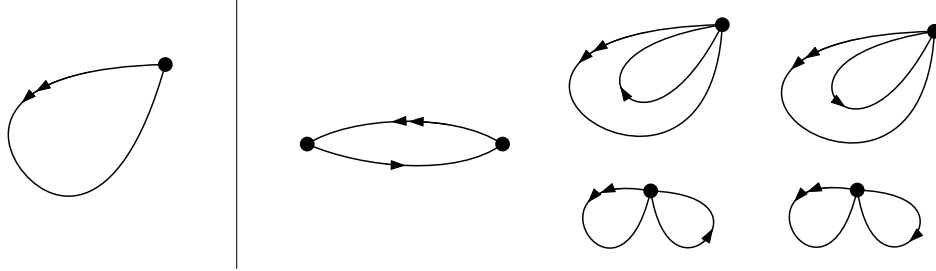


Figure 4.2: The planar Eulerian orientations with at most two edges, in agreement with  $G(t) = t + 5t^2 + O(t^3)$ . On the right are the four quartic Eulerian orientations with one vertex, in agreement with  $Q(t) = 4t + O(t^2)$ .

of such orientations having  $n$  vertices behaves asymptotically as

$$q_n \sim \kappa \frac{\mu^{n+2}}{n^2 (\log n)^2}$$

where

$$\kappa = 1/18 \quad \text{and} \quad \mu = 4\sqrt{3}\pi.$$

The series  $Q(t)$  is not  $D$ -finite, which means that it does not satisfy any non-trivial linear differential equation.

The first coefficients of  $R$  and  $Q$  are

$$R(t) = t - 3t^2 - 12t^3 - 105t^4 - 1206t^5 - \dots, \quad Q(t) = 4t + 35t^2 + 402t^3 + \dots.$$

### Remarks

1. As we will explain in Section 4.2.2, the series  $Q(t)$  also counts (by faces) quadrangulations equipped with a proper 3-colouring of the vertices (with prescribed colours on the root edge). It is worth noting that the generating functions of 3-coloured triangulations, and of 3-coloured planar maps, are both algebraic [23] (and thus  $D$ -finite), hence in a sense they are much simpler. The corresponding asymptotic estimates are  $\kappa\mu^n n^{-5/2}$  in both cases (for other values of  $\mu$  and  $\kappa$  of course).
2. In Section 4.7.2, we will prove that  $Q(t)$  also counts, by edges, Eulerian *partial* orientations of planar maps: that is, only some edges are oriented, with the condition that at any vertex there are as many incoming as outgoing edges.
3. As mentioned above, the series  $R(t)$  already occurs in the map literature, and more precisely in the enumeration of quartic maps  $M$  weighted by their Tutte polynomial  $T_M(0, 1)$ . However, our proof does not rely on this

observation, and it remains an open problem to understand this connection combinatorially. We refer to the final section for more detail.

The counterpart of Theorem 4.1 for all rooted planar Eulerian orientations reads as follows.

**Theorem 4.2.** *Let  $R(t) \equiv R$  be the unique formal power series with constant term 0 satisfying*

$$t = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n}^2 R^{n+1}.$$

*Then the generating function of rooted planar Eulerian orientations, counted by edges, is*

$$G(t) = \frac{1}{4t^2} (t - 2t^2 - R(t)).$$

*This is a differentially algebraic series, satisfying a non-linear differential equation of order 2 whose coefficients are polynomials in  $t$ . The number  $g_n$  of such orientations having  $n$  vertices behaves asymptotically as*

$$g_n \sim \kappa \frac{\mu^{n+2}}{n^2 (\log n)^2}$$

where

$$\kappa = 1/16 \quad \text{and} \quad \mu = 4\pi.$$

*The series  $G(t)$  is not D-finite.*

The first coefficients of  $R$  and  $G$  are

$$R(t) = t - 2t^2 - 4t^3 - 20t^4 - 132t^5 - \dots, \quad G(t) = t + 5t^2 + 33t^3 + \dots.$$

**Remark.** In Section 4.7 we will prove that  $2G(t)$  also has an interpretation in terms of 3-coloured maps: it counts (by faces) properly 3-coloured quadrangulations having no bicoloured face. Equivalently, it counts Eulerian orientations of quartic maps with no *alternating vertex* (a vertex where the order of the edges would be in/out/in/out). This is the special case  $\alpha = \beta$  of a two matrix model studied in [137], where the point  $\alpha = \beta = 1/(4\pi)$  is indeed identified as critical.

**Outline of the chapter.** In Section 4.2 we begin with basic definitions on maps, orientations, and generating functions. We also discuss various models related to quartic Eulerian orientations. In Section 4.3 we describe the numerical work from [91], in which we conjectured the asymptotic forms of the two generating functions  $Q(t)$  and  $G(t)$ . This begins in Sections 4.3.2

and 4.3.3 with a description of systems of functional equations which characterise the two generating functions. Using these equations, we compute 100 coefficients of  $Q(t)$  and 90 coefficients of  $G(t)$ . We present an empirical analysis of these terms in Section 4.3.5, which allows us to conjecture the exact asymptotic form of both generating functions. In Section 4.4, we describe how these asymptotic conjectures led us to conjecture the exact form of both  $Q(t)$  and  $G(t)$ . The proof of these conjectures begins in Section 4.5, where we write a simpler system of functional equations which characterises the generating function  $Q(t)$  of quartic Eulerian orientations. We solve it in Section 4.6, using a guess-and-check approach. Then comes a bijective *intermezzo* in Section 4.7, where we describe a bijection, designed by Bouttier, Fusy and Guitter [53]. As a special case, this bijection implies that general Eulerian orientations with  $n$  edges are in one-to-two correspondence with certain restricted quartic Eulerian orientations with  $n$  vertices. In Section 4.8 we give a system of equations for these orientations, which we solve in Section 4.9. In Section 4.10 we briefly discuss the nature of our generating functions and their singular behaviour, thus proving the asymptotic statements in Theorems 4.1 and 4.2, which are also the conjectures in Section 4.3.5. Section 4.11 finally raises some open problems.

## Definitions

### Planar maps

A *planar map* is a proper embedding of a connected planar graph in the oriented sphere, considered up to orientation preserving homeomorphism. Loops and multiple edges are allowed (Figure 4.3). The *faces* of a map are the connected components of its complement. The numbers of vertices, edges and faces of a planar map  $M$ , denoted by  $v(M)$ ,  $e(M)$  and  $f(M)$ , are related by Euler's relation  $v(M) + f(M) = e(M) + 2$ . The *degree* of a vertex or face is the number of edges incident to it, counted with multiplicity. A *corner* is a sector delimited by two consecutive edges around a vertex; hence a vertex or face of degree  $k$  is incident to  $k$  corners. The *dual* of a map  $M$ , denoted  $M^*$ , is the map obtained by placing a vertex of  $M^*$  in each face of  $M$  and an edge of  $M^*$  across each edge of  $M$ ; see Figure 4.3, right. A map is said to be *quartic* if every vertex has degree 4. Duality transforms quartic maps into *quadrangulations*, that is, maps in which every face has degree 4. A planar map is *Eulerian* if every vertex has even degree. Its dual, with even face degrees, is then *bipartite*. We call a face of degree 2 (resp. 4) a *digon* (resp. *quadrangle*).

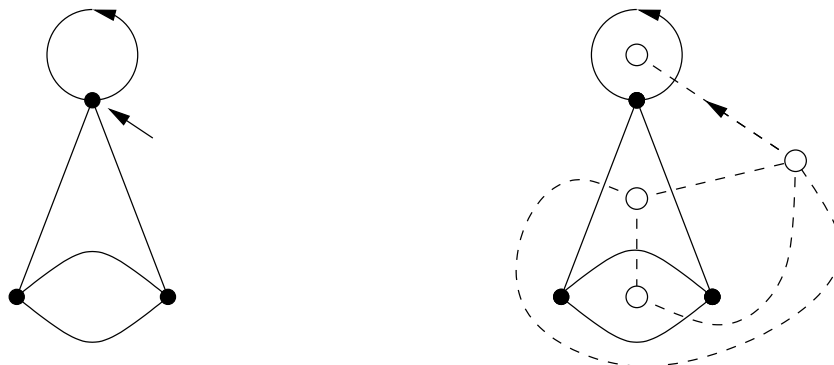


Figure 4.3: Left: a rooted planar map, with the root edge and root corner shown. Right: the dual map, in dashed edges.

For counting purposes it is convenient to consider *rooted* maps. A map is rooted by choosing an edge, called the *root edge*, and orienting it. The starting point of this oriented edge is then the *root vertex*, the other endpoint is the *co-root vertex*. The face to the right of the root edge is the *root face*, and its edges are the *outer edges*. The face to the left of the root edge is the *co-root face*. Equivalently, one can root the map by selecting a corner. The correspondence between these two rooting conventions is that the oriented root edge follows the root corner in anticlockwise order around the root vertex. In figures, we usually choose the root face as the infinite face (Figure 4.3). This explains why we often call the root face the *outer face* and its degree the *outer degree* (denoted  $\text{od}(M)$ ). The other faces are called *inner faces*. Similarly, we call the corners of the outer face *outer corners* and all other corners *inner corners*.

From now on, every map is *planar* and *rooted*, and these precisions will often be omitted. Our convention for rooting the dual of a map is illustrated on the right of Figure 4.3. Note that it makes duality of rooted maps a transformation of order 4 rather than 2. By convention, we include among rooted planar maps the *atomic map* having one vertex and no edge.

## Orientations

A (planar) *Eulerian orientation* is a (rooted, planar) map with oriented edges, in which the in- and out-degrees of each vertex are equal. We require that the orientation chosen for the root edge is consistent with its orientation coming from the rooting (Figure 4.1, right). Note that the underlying map must be Eulerian. We find it convenient to work with duals of Eulerian orientations, which turn out to be equivalent to certain *labelled maps*.

**Definition 4.3.** A *labelled map* is a rooted planar map with integer labels on its vertices, such that adjacent labels differ by  $\pm 1$  and the root edge is labelled from 0 to 1. Such a map is necessarily bipartite. We also consider the atomic map, with a single vertex (labelled 0), to be a labelled map.

An example is shown in Figure 4.4.

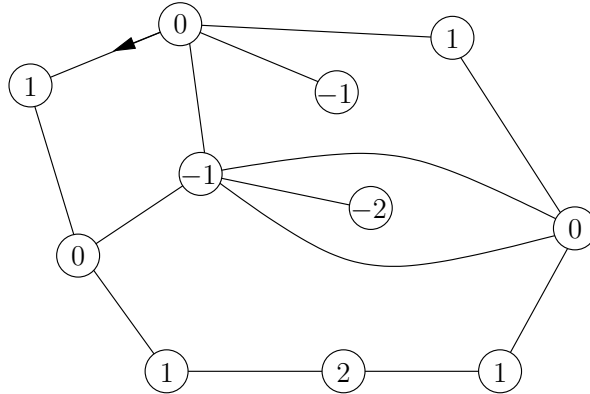


Figure 4.4: A labelled map.

**Lemma 4.4.** *The duality transformation between Eulerian maps and bipartite maps can be extended into a bijection between Eulerian orientations and labelled maps, which preserves the number of edges and exchanges vertex degrees and face degrees.*

*Proof.* Given a labelled map  $L$ , we can construct a directed map by orienting each edge from the lower number to the higher number. Then around each face of  $L$ , the number of clockwise edges is equal to the number of anticlockwise edges. Hence the dual of this map (where the orientations of the edges are defined by rotating the original edges  $90^\circ$  clockwise) is an Eulerian orientation. This transformation is illustrated in Figure 4.5. The restriction that the root edge of  $L$  is labelled from 0 to 1 is equivalent to the restriction that the root edge of an Eulerian orientation is directed away from the root vertex. By reversing each of these steps, we see that this transformation is a bijection. Hence, the number of labelled maps with  $n$  edges is equal to the number of rooted planar Eulerian orientations with  $n$  edges. Using the same bijection, we see that the number of 4-valent rooted planar Eulerian orientations with  $n$  edges is equal to the number of labelled quadrangulations with  $n$  edges.  $\square$

In the case of a quartic Eulerian orientation, the (quadrangular) faces of the dual map can only have two types of labelling, shown in Figure 4.6. It is

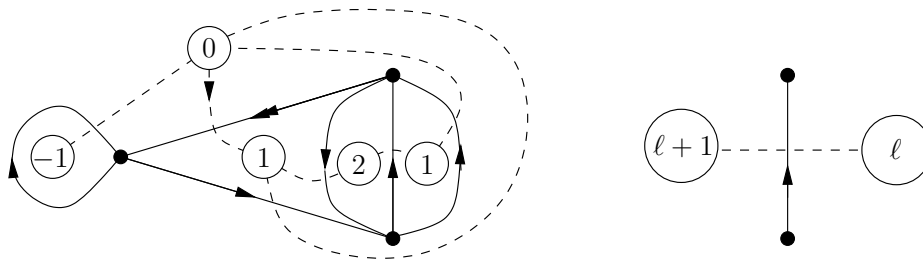


Figure 4.5: An Eulerian orientation (solid edges) and the corresponding dual labelled map (dashed edges). The labelling rule is shown on the right.

easily shown that, upon replacing every label by its value modulo 3, one obtains a proper 3-colouring of the vertices of the quadrangulation. Conversely, given a 3-coloured quadrangulation such that the root edge is oriented from 0 to 1, one can directly reconstruct an Eulerian orientation of the dual quartic map using the rule of Figure 4.5. Then the associated labelled quadrangulation projects on the coloured quadrangulation modulo 3. Hence 4-valent Eulerian orientations with  $n$  vertices are in bijection with 3-coloured quadrangulations with  $n$  faces (with the root edge oriented from 0 to 1), as claimed below Theorem 4.1. This correspondence between Eulerian orientations of a planar quartic graph and 3-colourings of its dual has been known for a long time. In the more general, non-planar case, the number of Eulerian orientations of a 4-valent graph is given by the value of its *Tutte polynomial* at the point  $(0, -2)$ , with no interpretation in terms of colourings [211, Sec. 3.6].

**More orientations.** Obviously, quartic Eulerian orientations are orientations of a quartic map with exactly 2 outgoing edges at each vertex. It turns out that the number of oriented *quadrangulations* in which each vertex has outdegree 2 is known. The associated series is D-finite. A simple bijection transforms these orientations into *bipolar* orientations of planar maps (no cycle, one source, one sink, both on the outer face). We refer the reader to [17, 43, 102, 111], and references therein. Analogous results exist for orientations of triangulations in which every vertex has outdegree 3, called *Schnyder orientations* [22, 41]. Let us also mention recent progress regarding bipolar orientations with prescribed face degrees [50].

## Eulerian orientations, the 6-vertex model and fully packed loops

The enumeration of quartic Eulerian orientations has already been considered in the mathematical physics literature, where it is called the ice model on a 4-valent random lattice [143, 214]: an oxygen atom stands at every vertex,

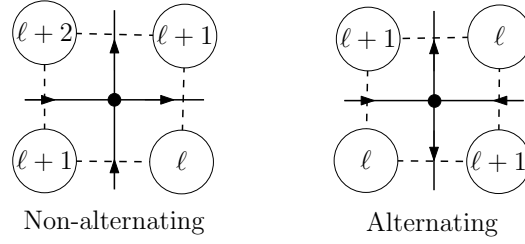


Figure 4.6: The two types of vertices in a quartic Eulerian orientation, shown with the associated quadrangles in the dual labelled map. In the six vertex model, alternating vertices are assigned the weight  $\omega$ .

while the hydrogen atoms (two per oxygen in a water/ice molecule) lie on the edges, the arrows indicating with which oxygen they go. This is a special case of the *six vertex model*: in that model, the configurations are still Eulerian orientations, but a weight  $\omega = 2 \cos(\lambda\pi/2)$  is assigned to each vertex from which the two outgoing edges are opposite each other. We call these vertices *alternating* (see Figure 4.6, right). The ice model then corresponds to  $\omega = 1$ , or equivalently,  $\lambda = 2/3$ . Starting from a matrix integral formulation, Kostov [143] and Zinn-Justin [214] predict that for  $\lambda \in [0, 2)$ , that is,  $\omega \in (-2, 2]$ , the dominant singularity occurs at

$$\rho = \frac{1}{8\lambda\pi} \frac{\sin(\lambda\pi/4)}{\cos(\lambda\pi/4)^3}. \quad (4.1)$$

In terms of  $\omega$ , this is

$$\rho = \frac{1}{4 \arccos(\omega/2)} \frac{\sqrt{2-\omega}}{(2+\omega)^{3/2}}.$$

Moreover, Kostov [143] predicts that the behaviour of the free energy around this singularity is

$$\log(Z(t, \omega)) \sim \frac{(1 - t/\rho)^2}{\log(1 - t/\rho)},$$

up to some multiplicative constant.

In a forthcoming paper we (Bousquet-Mélou and I) consider the refined generating function  $\mathbf{Q}(t, \omega)$  of quartic Eulerian orientations, where  $t$  still counts vertices, and a weight  $\omega$  is assigned to alternating vertices. Figure 4.6 shows that  $\mathbf{Q}(t, \omega)$  also counts labelled quadrangulations by faces, with a weight  $\omega$  per face having only two distinct labels. In terms of the free energy considered by Kostov, our generating function  $\mathbf{Q}(t, \omega)$  is given by

$$\mathbf{Q}(t, \omega) = 2t \frac{d}{dt} \log(Z(t, \omega)),$$

where the derivative accounts for the  $2n$  possible choices of the root edge in a quartic Eulerian orientation with  $n$  vertices. Its predicted singular behaviour is thus

$$Q(t, \omega) \sim \frac{1 - t/\rho}{\log(1 - t/\rho)},$$

up to some multiplicative constant.

The generating function  $Q(t)$  of Theorem 4.1 is  $Q(t, 1)$ , so the predictions of Zinn-Justin and Kostov at  $\omega = 1$  are verified by Theorem 4.1 (see Proposition 4.37 for the singular behaviour of  $Q(t)$ , from which the asymptotic behaviour of the numbers  $q_n$  stems).

But we also claim that our second theorem, Theorem 4.2, solves the case  $\omega = 0$  of the six vertex model. Indeed, we will show that the generating function  $G(t)$  of general Eulerian orientations satisfies  $2G(t) = Q(t, 0)$  (see Corollary 4.29). Hence the predictions of Zinn-Justin and Kostov for  $\omega = 0$  follow from Theorem 4.2 and Proposition 4.38.

### Remarks

1. The case  $\omega = 2$  is also well-understood, and boils down to counting all planar maps weighted by their *Tutte polynomial*, evaluated at the point  $(3, 3)$ . This can be justified as follows: starting from a quartic Eulerian orientation, we first transform each vertex into a pair of vertices with degree 3, as shown in Figure 4.7. (This transformation has already been used in, e.g., [143, 214]). The two possible choices for each alternating vertex account for the weight  $\omega = 2$  assigned to these vertices. The resulting map is a cubic map in which each vertex has one incoming edge, one outgoing edge and one undirected edge. This transformation can be reversed by simply contracting all undirected edges in the cubic map. The oriented edges must then form loops on the cubic map, where each loop is oriented one of two ways - either clockwise or anticlockwise. Moreover, every vertex must be visited by a loop. In [45, Sec. 2.1], this model of fully packed loops on cubic maps is shown to be equivalent to the 4-state Potts model on general planar maps, in which every monochromatic edge gets a weight 3, and every vertex a weight  $1/2$ . Finally, using the correspondence between the Potts model and the Tutte polynomial (see, e.g., [23, Sec. 3.3]), we conclude that

$$Q(t, 2) = \sum_{M \text{ planar}} t^{e(M)} T_M(3, 3) = 6t + 78t^2 + 1326t^3 + 25992t^4 + O(t^5),$$

where  $T_M(\mu, \nu)$  is the Tutte polynomial of  $M$  (see [211]). This series was proved to satisfy an (explicit) non-linear differential equation of degree 3 (see [24, Thm. 16] for  $\beta = 2$ ), but, to our knowledge, the singular behaviour of  $Q(t, 2)$  has not been derived from it. One can in fact guess-and-prove a



smaller differential equation, of order 2.

**2.** In our forthcoming paper, we give a system of equations defining  $Q(t, \omega)$ . The coefficients that we thus compute suggest that the prediction (4.1) does not hold on the entire segment  $(-2, 2]$ , but only for  $\omega > \omega_c$ , where  $\omega_c$  is around  $-0.76$ .

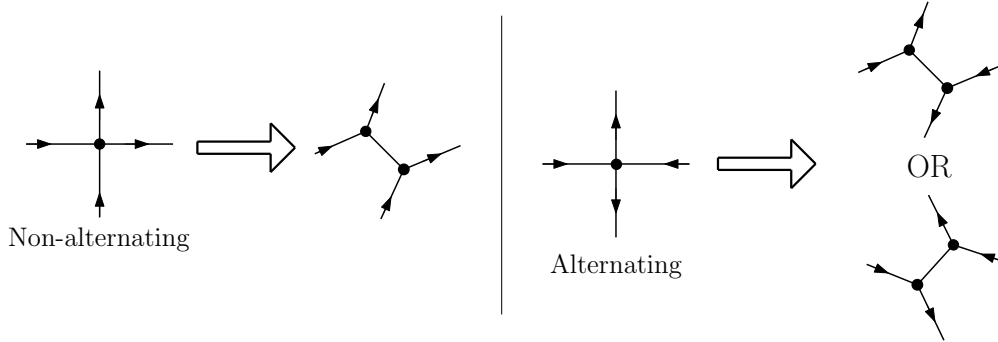


Figure 4.7: The transformation from vertices in a quartic Eulerian orientation to pairs of vertices in certain cubic Eulerian partial orientations. This transformation applies when  $\omega = 2$ .

## Numerical work

In this section we describe the numerical work which allowed us to conjecture the asymptotic form of the generating functions  $Q(t)$ , for quartic Eulerian orientations counted by vertices, and  $G(t)$  for general Eulerian orientations counted by edges. For more detail on this work see [91]. This is also what led us to conjecture the exact forms of the solutions, which we subsequently proved. The conjectures in this Section are all proven in later sections of this chapter, so the impatient reader may skip to Section 4.5 for the beginning of the solution.

### Contraction Operation

We will now describe our contraction operation, which is a fundamental ingredient in our derivations of systems of functional equations which characterise the generating functions  $G(t)$  and  $Q(t)$ . This definition is illustrated in Figure 4.28. Similar contraction operations will be used in Sections 4.5 and 4.8 when we derive simpler systems of functional equations.

**Definition 4.5.** Let  $L$  be a labelled map and let  $c$  be an outer corner of  $L$  labelled 0. We define the *minus-subpatch* of  $L$  rooted at  $c$  as follows. First,

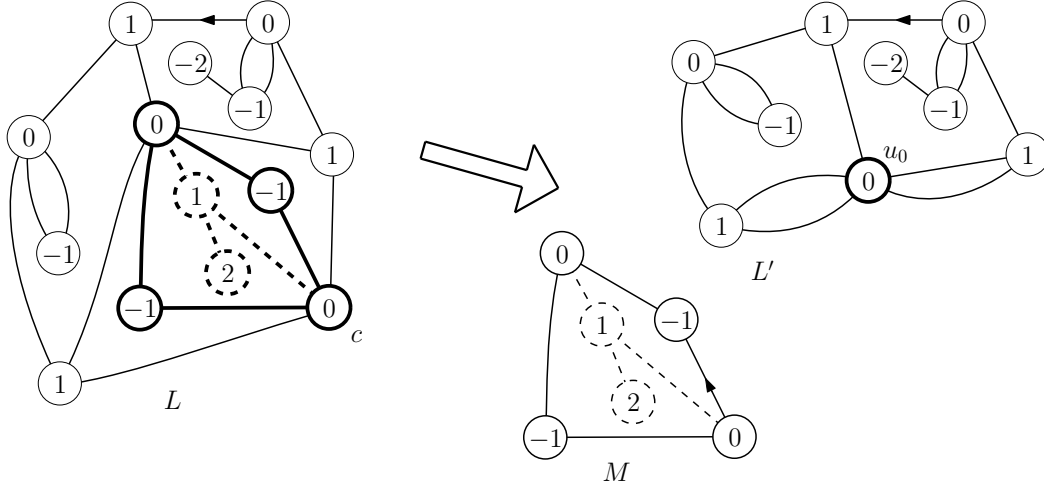


Figure 4.8: Left: a labelled map  $L$  with a chosen outer corner  $c$  labelled 0. The minus-subpatch  $M$  of  $D$  rooted at  $c$  is highlighted in  $D$  and shown separately in the middle. The submap  $M'$  is obtained from  $M$  by deleting the dashed vertices and edges. Right: the labelled map  $L'$  constructed from  $L$  by contracting  $M$  to a single vertex  $u_0$ .

let  $M'$  be the maximal submap of  $L$  that contains  $v$  and consists of vertices labelled 0 or less. Let  $M$  be the submap of  $L$  that contains  $M'$  and all edges and vertices within its boundary (assuming the root face is drawn as the infinite face). The map  $M$ , which we root at the corner inherited from  $c$ , is the *minus-subpatch* of  $L$  rooted at  $c$ .

It follows immediately from this definition that  $M$  and  $M'$  share the same outer face. Moreover, all inner faces of  $M$  are also inner faces of  $L$ . All outer vertices of  $M$  must also be outer vertices of  $M'$ , so they have non-positive labels. In particular, this implies that the co-root vertex of  $L$  is not a vertex of  $M'$ . Note also that since the root vertex of  $M$  is labelled 0, the co-root vertex of  $M$  is labelled  $-1$ , so  $M$  is not a labelled map but we can create a labelled map  $\hat{M}$  by replacing each label  $\ell$  in  $M$  with  $-\ell$ .

Now, every edge in  $L$  which connects a vertex in  $M$  to a vertex not in  $M$  must have endpoints labelled 0 (in  $M$ ) and 1 (not in  $M$ ). We can thus form a new labelled map  $L'$  by contracting all of  $M$  to a single vertex  $u_0$  labelled 0 (Figure 4.28). This vertex  $u_0$  is only adjacent to vertices labelled 1 in  $L'$ . We define the root edge of  $L'$  to be the edge corresponding to the root edge of  $L$ , and we call  $L'$  the  *$M$  – contraction* of  $L$ .

Thus given a labelled map  $L$ , and an outer corner  $c$  labelled 0, we can decompose  $L$  into two labelled maps  $L'$  and  $\hat{M}$ . In general this transformation is not uniquely reversible, and there are restrictions on the possible maps  $L'$

and  $\hat{M}$ . Nonetheless, we will use this transformation in certain restricted situations where it is a bijection, and this will be a key step in our solution to the problem of enumerating quartic Eulerian orientations.

## Functional equations for quartic Eulerian orientations

In this section we will characterise the generating function  $Q(t)$  of labelled quadrangulations by a system of functional equations.

**Theorem 4.6.** *There exists a unique pair of series, denoted  $P(t, x, y)$  and  $D(t, w, x, y)$ , belonging respectively to  $\mathbb{Q}[x, y][[t]]$  and  $\mathbb{Q}[w, x, y][[t]]$  satisfying the following equations:*

$$\begin{aligned} P(t, x, y) &= 1 + [w^0]D(t, w, x, y)P\left(t, 1, \frac{1}{w}\right), \\ D(t, w, x, y) &= t^2 x^2 \frac{D(t, w, x, y) - D(t, w, 1, y)}{x - 1} \\ &\quad + txD(t, w, x, y)(w + 2[y^1]P(t, x, y)) \\ &\quad + txy(1 + D(t, w, 1, y))P(t, x, y). \end{aligned}$$

The generating function  $Q(t)$  that counts labelled quadrangulations by faces is given by

$$Q(t^2) = \frac{1}{t}[y^1]P(t, 1, y) - 1.$$

By Lemma 4.4, the series  $Q(t)$  also counts quartic Eulerian orientations by vertices.

It is easy to see that this pair of equations defines  $P$  and  $D$  uniquely as follows: From the equation for  $D$ , one can deduce that  $D(0, w, x, y) = 0$ , then the first equation implies that  $P(0, x, y) = 1$ . Now, if  $n$  is a positive integer and the coefficients  $[t^k]P(t, x, y)$  and  $[t^k]D(t, x, y)$  are known for  $k < n$ , then the second equation determines  $[t^n]D(t, w, x, y)$ , then the first equation determines  $[t^n]P(t, x, y)$  (using the fact that  $D(0, w, x, y) = 0$ ). Hence, by induction,  $[t^n]P(t, x, y)$  and  $[t^n]D(t, x, y)$  can be determined for all  $n$ . Moreover, the series constructed inductively in this way will satisfy the pair of equations in the theorem. This inductive approach is also the basis of our algorithm that computes the coefficients of  $Q(t)$ , which are analysed in Section 4.3.5.

The series  $D$  and  $P$  of Theorem 4.6 count certain labelled maps, which we now define. See Figure 4.9 for an illustration.

**Definition 4.7.** A *patch* is a labelled map in which each inner face has degree 4, and the vertices around the outer face are alternately labelled 0 and 1.

*D-patches* resemble patches but may include digons. More precisely, a D-patch is a labelled map in which each inner face has degree 2 or 4, those of degree 2 being incident to the root vertex, and the vertices around the outer face are alternately labelled 0 and 1. We also require that all neighbours of the root vertex are labelled 1.

We define  $P(t, x, y)$  and  $D(t, w, x, y)$  to be the generating functions of patches and non-atomic D-patches, respectively where  $t$  counts edges,  $y$  the outer degree (halved),  $x$  the degree of the co-root vertex and  $w$  the number of inner digons. The atomic map has no co-root vertex, but we use the convention that this map contributes 1 to  $P(t, x, y)$ .

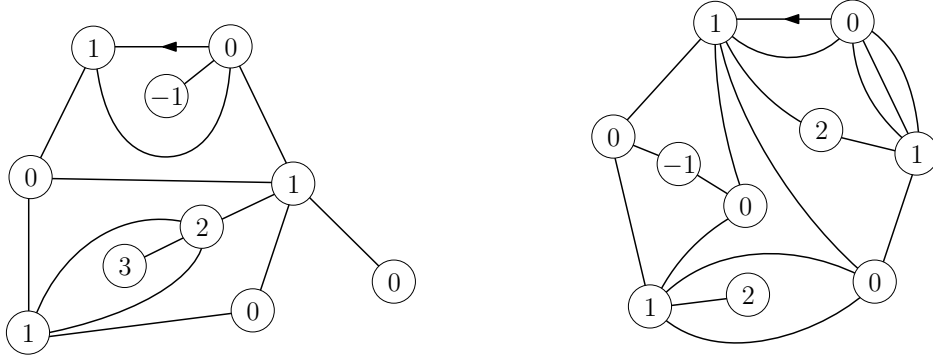


Figure 4.9: Left: a patch which contributes  $t^5 x^3 y^4$  to the generating function  $P(t, x, y)$ . Right: a D-patch which contributes  $t^6 w^3 x^6 y^3$  to the generating function  $D(t, w, x, y)$ .

It remains to show that the series  $P(t, x, y)$ ,  $D(t, w, x, y)$  and  $Q(t)$  satisfy the three equations of Theorem 4.6. We will now outline the proofs of these equations. For full details see [91].

**Lemma 4.8.** *The generating functions  $Q$  and  $P$  satisfy the equation*

$$Q(t^2) = \frac{1}{t} [y^1] P(t, 1, y) - 1,$$

*Proof.* For each positive integer  $k$ , the coefficient of  $t^k$  on the right hand side of the equation is the number of patches with outer degree 2 containing  $k + 1$  edges. Removing the outer edge other than the root edge from such a patch yields a labelled quadrangulation with  $k$  edges, and hence  $k/2$  faces, which also proves that for a patch to exist,  $k$  must be even. These are exactly the objects counted by the coefficient of  $t^k$  on the left hand side, hence

$$[t^k] Q(t^2) = [t^k] \left( \frac{1}{t} [y^1] P(t, 1, y) - 1 \right)$$

for  $k > 0$ . For  $k \leq 0$ , this equation still holds as both sides are equal to 0. Since this holds for all  $k$ , the lemma is true.  $\square$

**Lemma 4.9.** *The generating functions  $P$  and  $D$  satisfy the equation*

$$P(t, x, y) = 1 + [w^0]D(t, w, x, y)P\left(t, 1, \frac{1}{w}\right).$$

This is Proposition 2.5 in [91], and a full proof is given there. The idea is that a patch  $P$  which is not the atomic patch can be decomposed into a D-patch and a patch as follows: Let  $M$  be the minus subpatch of  $P$  rooted at the root corner, let  $D$  be the  $M$ -contraction of  $P$  and let  $\hat{P}$  be the labelled map obtained by replacing each label  $\ell$  in  $M$  with  $-\ell$ . Then one can show that  $D$  is a D-patch,  $P$  is a patch and the number of inner digons in  $D$  is equal to half the outer degree of  $\hat{P}$ . Moreover, subject to this restriction, one can uniquely reconstruct  $P$  from any pair  $D, \hat{P}$ . The equation follows by calculating the number of edges, the degree of the co-root vertex and the degree of the outer face of  $P$  in terms of those same statistics for  $\hat{P}$  and  $D$ .

**Lemma 4.10.** *The generating functions  $P$  and  $D$  satisfy the equation*

$$\begin{aligned} D(t, w, x, y) = & t^2 x^2 \frac{D(t, w, x, y) - D(t, w, 1, y)}{x - 1} \\ & + tx D(t, w, x, y)(w + 2[y^1]P(t, x, y)) \\ & + txy(1 + D(t, w, 1, y))P(t, x, y). \end{aligned}$$

This is Proposition 2.4 in [91], and a full proof is given there. We give an outline here:

The equation follows from investigating all cases which can occur when the root edge  $e$  is removed from a D-patch  $D$ . The contribution from the cases in which  $e$  is a bridge (that is, its removal disconnects the graph) is

$$txy(1 + D(t, w, 1, y))P(t, x, y).$$

From the cases in which the face  $F$  on the left of  $e$  is an inner digon, the contribution is

$$twx D(t, w, x, y).$$

From the cases in which  $F$  is an inner quadrangle but the two vertices incident to  $F$  labelled 1 are in fact the same vertex, the contribution is

$$2tx D(t, w, x, y)[y^1]P(t, x, y).$$

The contribution from the remaining cases is

$$t^2 x^2 \frac{D(t, w, x, y) - D(t, w, 1, y)}{x - 1}.$$

To understand this final contribution, let  $v_0$ ,  $v_1$ ,  $u$  and  $u_1$  be the vertices around  $F$  in anticlockwise order, with  $v_0$  and  $v_1$  the root vertex and co-root vertex of  $D$ . Let  $D'$  be the labelled map formed from  $D$  by moving  $u_1$  towards  $v_1$  inside  $F$  until they are identified, thereby forming a digon between  $v_0$  and  $v_1$  and another digon between  $v_1$  and  $u$ . Next, let  $D''$  be the map formed by replacing each of these two digons with a single edge. Then  $D''$  is a D-patch. To reconstruct  $D'$  from  $D''$ , we simply need to choose one edge incident to the co-root vertex to split into two edges, and then split the root edge into two edges, thereby forming two digons.  $D$  can then be reconstructed uniquely by splitting the co-root vertex into two vertices  $u_1$  and  $v_1$  in such a way that the two digons merge to form a single quadrangle  $F$ . Tracking the relevant statistics, it follows that D-patches  $D$  corresponding to a D-patch  $D''$  with  $n$  edges,  $m$  digons, an outer face with degree  $2d$  and a co-root vertex of degree  $k$  contribute

$$t^{n+2} w^m y^d (x^{k+1} + x^k + \dots + x^2).$$

Summing this over all D-patches  $D'$  yields the contribution

$$t^2 x^2 \frac{D(t, w, x, y) - D(t, w, 1, y)}{x - 1}$$

from this case.

## Functional equations for general Eulerian orientations

In this section we will characterise the generating function  $G(t)$  of labelled maps by a system of functional equations. For a simpler system of equations characterising  $G(t)$ , see Section 4.8.

**Theorem 4.11.** *There exist unique series, denoted  $P(t, y)$ ,  $R(t, w, x)$ ,  $S(t, w, x)$ ,  $T(t, w, x, y)$ ,  $U(t, x, y)$  and  $V(t, w, y)$ , belonging respectively to  $\mathbb{Q}[y][[t]]$ ,  $\mathbb{Q}[w, x][[t]]$ ,  $\mathbb{Q}[w, x][[t]]$ ,  $\mathbb{Q}[w, x, y][[t]]$ ,  $\mathbb{Q}[x, y][[t]]$  and  $\mathbb{Q}[w, y][[t]]$  satisfying the following*

equations:

$$\begin{aligned}
P(t, y) &= \Omega_x(V(t, x, y)), \\
R(t, w, x) &= wxt + \frac{1}{wxt}R(t, w, x)S(t, w, x), \\
S(t, w, x) &= \Omega_z \left( t^2w^2x^2 + \frac{zS(t, w, x) - xS(t, w, z)}{z(x - z)}R(t, w, z)b + \frac{w^2}{z^2}R(t, z, x)S(t, z, x) \right), \\
T(t, w, x, y) &= \Omega_z \left( \frac{T(t, w, x, y) - T(t, w, z, y)}{x - z}R(t, w, z)x \right) \\
&\quad + tx(y - w)V(t, x, y)U(t, w, y) + twxV(t, x, y), \\
U(t, w, y) &= \Omega_x \left( \frac{1}{twx}T(t, w, x, y)R(t, w, x) \right) + 1, \\
V(t, x, y) &= \Omega_w \left( \frac{1}{twx}T(t, w, x, y)R(t, w, x) \right) + 1,
\end{aligned}$$

along with the initial conditions  $R(0, w, x) = S(0, w, x) = T(0, w, x, y) = 0$ , where  $\Omega_z$  is a linear operator on series in  $z$  defined by

$$\Omega_z \left( \sum_{n=0}^{\infty} c_n z^n \right) = c_0 + \sum_{n=1}^{\infty} \sum_{j=0}^{\infty} c_n \binom{n+j-1}{n-1} [y^j] P(t, y),$$

where each term  $c_n$  is a series in variables other than  $z$ .  $\Omega_x$  and  $\Omega_w$  are defined similarly. The generating function  $G(t)$  that counts labelled quadrangulations by faces is given by

$$G(t) = \Omega_w \left( \Omega_x \left( \frac{1}{t^2w^2x^2} R(t, w, x) S(t, w, x) \right) \right).$$

By Lemma 4.4, the series  $G(t)$  also counts Eulerian orientations by edges.

The proof of this theorem is in [91], but we will only describe two important elements of the proof here. The first is the proof that this equation uniquely defines the series involved.

**Lemma 4.12.** *The system of equations in Theorem 4.11 uniquely defines the series involved.*

*Proof.* First, we observe that if  $N$  is a positive integer and  $A(t, z)$  is some series in  $t, z$  and possibly other indeterminates such that  $A(0, z) = 0$  and we know  $[t^k]P(t, y)$  for  $k < N$  and  $[t^k]A(t, z)$  for  $k \leq N$ , then this determines the coefficients  $[t^k]\Omega_z(A(t, z))$  for  $k \leq N$ , using the equation defining  $\Omega_z$ .

Now, it follows from the third equation in the Theorem that  $S(t, w, x)$  is divisible by  $t^2$ , then using this equation again, it follows that  $[t^2]S(t, w, x) = w^2x^2$ . The first, fifth and sixth equations allow us to determine that  $V(0, x, y) = U(0, x, y) = P(0, x, y) = 1$ . Now we will proceed by induction. Assume that the coefficient of  $t^k$  can be determined for each series for  $k < N$ , where  $N$  is some positive integer. Then the coefficients  $[t^N]S(t, w, x)$  and  $[t^N]T(t, w, x, y)$  are determined by the third and fourth equations, respectively. Next, the coefficient  $[t^N]R(t, w, x)$  is determined by the second equation. Then the coefficients  $[t^N]U(t, w, x)$  and  $[t^N]V(t, w, x)$  are determined by the fifth and sixth equations. Finally, rewriting the first equation as

$$P(t, y) = \Omega_x(V(t, x, y) - 1) + 1,$$

we see that the coefficient  $[t^N]P(t, y)$  can also be determined. Hence, by induction, the equations uniquely determine the six series involved. This also determines the series  $G(t)$ , using the final equation in the Theorem. Moreover, due to the way in which these series were constructed, it is clear that they satisfy the equations. Hence there are unique series  $P, R, S, T, U, V$  which satisfy the equations and initial conditions.  $\square$

The series  $P, R, S, T, U$  and  $V$  of Theorem 4.6 count certain labelled maps, which we now define.

**Definition 4.13.** A *general patch* is a labelled map in which the vertices around the outer face have nonnegative labels.

A *U-patch* is a general patch in which the root vertex is only adjacent to vertices labelled 1.

A *V-patch* is a general patch in which the co-root vertex is only adjacent to vertices labelled 0.

A *T-patch* is a general patch which is a U-patch and a V-patch in which the only edge between the root vertex and co-root vertex is the root edge.

An *R-patch* is a general patch which is both a U-patch and a V-patch in which the outer face has degree 2.

An *S-patch* is an R-patch in which the root vertex and the co-root vertex are joined by exactly two edges (which must be the two edges of the outer face).

By convention, the atomic map is a general patch, a U-patch and a V-patch but not an R-patch, S-patch or T-patch.

We define  $P(t, y)$ ,  $R(t, w, x)$ ,  $S(t, w, x)$ ,  $T(t, w, x, y)$ ,  $U(t, w, y)$  and  $V(t, x, y)$  to be the generating functions of general patches, R-patches, S-patches, T-patches, U-patches and V-patches, respectively, where  $t$  counts edges,  $w$  the



degree of the root vertex,  $x$  the degree of the co-root vertex and  $y$  the number of corners of the outer face labelled 0.

It remains to show that these six series and  $\mathbf{G}(t)$  satisfy the equations of Theorem 4.6. We will show one element of the proof here, namely the interpretation of the linear operator  $\Omega_z$ . For the complete proof of these equations see [91].

**Proposition 4.14.** *Consider tuples  $P, c, M, P', c'$  of a general patch  $P$ , outer corner  $c$  of  $P$  labelled 0, minus-subpatch  $M$  of  $P$  rooted at  $c$ ,  $M$ -contraction  $P'$  of  $P$  and corner  $c'$  in  $P'$  inherited from  $c$ . For a given pair of a patch  $P'$  and outer corner  $c'$  labelled 0 for which all adjacent vertices are labelled 1, the generating function  $\mathbf{E}_{(P', c')}(t)$  for tuples  $P, c, M$  corresponding to  $P', c'$ , counted by edges in  $M$  is given by*

$$\mathbf{E}_{(P', c')}(t) = \Omega_z(z^n),$$

where  $n$  is the number of corners of  $P'$  which are incident to the same vertex as  $c'$ , but are not incident to the outer face.

*Proof.* Using the equation for  $\Omega_z$ , we have

$$\Omega_z(z^n) = \sum_{j=0}^{\infty} \binom{n+j-1}{n-1} [y^j] \mathbf{P}(t, y)$$

for  $n > 0$  and  $\Omega_z(z^0) = 1$ .

By the definition of a minus subpatch, all outer vertices of  $M$  must have non-positive labels. Hence, the map  $\hat{M}$  formed by replacing each label  $\ell$  in  $M$  with  $-\ell$  is a general patch. If  $n = 0$ , then  $M$  must be the atomic map. Hence in this case

$$\mathbf{E}_{(P', c')}(t) = 1 = \Omega_z(z^n).$$

Now we will consider the case when  $n \geq 1$ . First, highlight one of the edges in  $P'$  which is incident on the vertex  $v$  incident on  $c$ . Recall that general patches are enumerated by  $\mathbf{P}(t, y)$ . Consider a specific general patch  $\hat{P}$ , which contributes  $t^k y^j$  to  $\mathbf{P}(t, y)$ . Then  $\hat{P}$  contains  $k$  edges and its outer face is incident to  $j$  corners at height 0. We will calculate the contribution to  $\mathbf{E}_{(P', c')}(t)$  of all cases in which  $\hat{M} = \hat{P}$ . Clearly any specific tuple  $P, c, M$  contributes  $x^k$  to  $M_v(x)$ , so we just need to calculate the number of these tuples. Assume that  $\hat{M} = \hat{P}$ . Then  $M$  is constructed from  $\hat{P}$  by replacing each label  $\ell$  in  $\hat{P}$  by  $-\ell$ . Going clockwise around the outer face of  $M$ , starting at the root vertex  $v_0$ , let  $p_1, p_2, \dots, p_j$  be the paths between vertices at height 0, so these partition the boundary of  $M$ . Now let  $c_1, c_2, \dots, c_n$  be the inner corners around  $v$  in  $P'$ , in clockwise order starting from the corner  $c$ . Then

in the transformation from  $P'$  to  $P$ , each inner corner  $c_i$  of  $P'$  expands to contain a number  $a_i$  of the paths  $p_1, \dots, p_j$ . Since the outer vertices of  $M$  which are not at height 0 have negative heights, each path  $p_t$  must not be on the outside of  $\Gamma$ , so  $p_t$  must be counted by one of the terms  $a_i$ . Moreover, due to the clockwise order, the lower expansion is uniquely determined by the sequence  $a_1, \dots, a_n$ , the only restrictions on this sequence being that each term  $a_i$  is a non-negative integer and the sum of the terms is  $j$ . The number of such sequences is

$$\binom{n+j-1}{n-1}.$$

Hence the contribution to  $E_{(P',c')}(t)$  of all pairs  $P, c$  corresponding to the general patch  $\hat{P}$  is

$$\binom{n+j-1}{n-1} t^k.$$

Summing this over all general patches  $\hat{P}$  yields the desired result:

$$E_{(P',c')}(t) = \sum_{j=0}^{\infty} \binom{n+j-1}{n-1} [y^j] P(t, y) = \Omega_z(z^n).$$

□

## Algorithmic computation of generating functions

As we have seen, we can use our functional equations to compute all of the involved series one coefficient (of  $t^n$ ) at a time. Since there are only polynomially many non-zero coefficients with degree in  $n$  at most  $t$  and each computation of a single coefficient takes polynomial time, this algorithm takes polynomial time. The coefficients were calculated *modulo* a prime smaller than  $2^{31}$ , repeated for several different primes, sufficient to calculate the coefficient by use of the Chinese Remainder Theorem. In this way we calculated 90 terms of the generating function  $G(t)$  for planar Eulerian orientations counted by edges and 100 terms of the generating function  $Q(t)$  for 4-valent planar Eulerian orientations counted by vertices.

## Empirical analysis of generating functions

We first tried to analyse the series  $Q(t)$  and  $G(t)$  by the method of differential approximants (DAs) [126, 122, 125]. The results were not totally straightforward. Assuming a power-law singularity of the form

$$f(t) \sim C(1 - t/t_c)^\alpha,$$

then for  $G(t)$  we found the closest singularity to the origin to be at  $t_c \approx 0.07957736$ , with an exponent around  $\alpha \approx 1.24$ . However there was a second singularity very close by, at  $t \approx 0.0795789$ , with an exponent around 2.26, and a third, less precisely located singularity at around  $t \approx 0.0798$ , with a complex exponent the value of which is irrelevant.

For  $Q(t)$  we found essentially identical results, just with a changed radius of convergence. In particular we found the closest singularity to the origin to be at  $t_c \approx 0.04594404$ , with an exponent around 1.23. There was a second singularity very close by, at  $t \approx 0.04594449$ , with an exponent around 2.23, and a third, less precisely located, at around  $t \approx 0.0459$ , with a complex exponent the value of which is irrelevant. This behaviour, where one has two singularities very close together, with an exponent separated by about 1.0, is known to be characteristic of a confluent singularity, and more precisely, a confluent singularity involving a logarithmic term. For  $Q(t)$ , this is consistent with Kostov's prediction [143] that

$$Q(t) \sim f(t) = \frac{-t(1 - \mu t)}{\log(1 - \mu t)}.$$

This also suggests that our estimates for of the critical exponents were inaccurate, which is not surprising given the presence of the logarithmic term.

Having seen that the method of DAs has difficulty in estimating the critical exponent, we turned to ratio-based methods. If the generating function  $f(t)$  behaves as above, then the coefficients behave as

$$[t^n]f(t) = \frac{c \cdot \mu^n}{n^2} \left( \frac{1}{\log^2 n} + \frac{a}{\log^3 n} + \frac{b}{\log^4 n} + \frac{c}{\log^5 n} + o\left(\frac{1}{\log^5 n}\right) \right). \quad (4.2)$$

To extract asymptotics from numerical data is difficult when successive terms are only weaker by a factor of a logarithm, which varies but slowly unless one has a vast number of terms.

The ratio of successive coefficients in this case behaves as

$$r_n = \frac{[t^n]f(t)}{[t^{n-1}]f(t)} = \mu \left( 1 - \frac{2}{n} - \frac{2}{n \log n} \left( 1 + \frac{c_1}{\log n} + \frac{c_2}{\log^2 n} + \frac{c_3}{\log^3 n} \right) + o\left(\frac{1}{n \log^4 n}\right) \right).$$

We show in figures 4.10 and 4.11 the ratios for  $G(t)$  and  $Q(t)$  plotted against  $1/n$ . Both plots exhibit slight concavity, due to the logarithmic corrections.

If we eliminate the  $O(1/n)$  term by constructing linear intercepts,

$$l_n = n \cdot r_n - (n-1) \cdot r_{n-1} = \mu \left( 1 + \frac{2}{n \log^2 n} + \frac{4c_1}{n \log^3 n} + \frac{6c_2}{n \log^4 n} + o\left(\frac{1}{n \log^4 n}\right) \right) \quad (4.3)$$

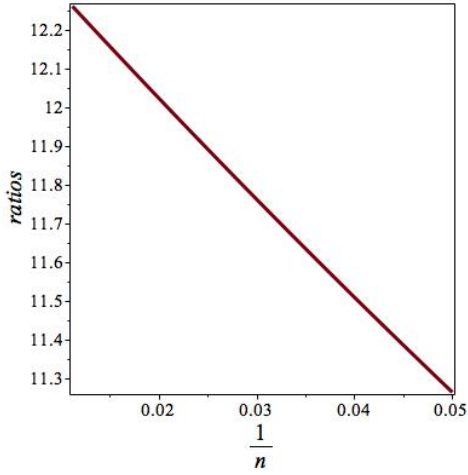


Figure 4.10: Ratio plot of coefficients of  $G(t)$ .

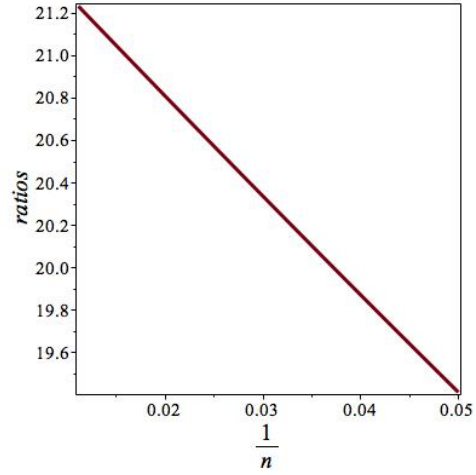


Figure 4.11: Ratio plot of coefficients of  $Q(t)$ .

the corresponding plots of the linear intercepts against  $1/(n \cdot \log^2 n)$  are shown in figures 4.12 and 4.13. Note that the ordinate is compressed by about a factor of 10, and secondly, the plot exhibits *more* curvature, presumably reflecting competition between subdominant logarithmic terms. Indeed, from the asymptotics, it is clear that this sequence must eventually have a positive gradient as  $n$  increases, so must pass through a maximum. We will see below that this occurs for sufficiently large  $n$ .

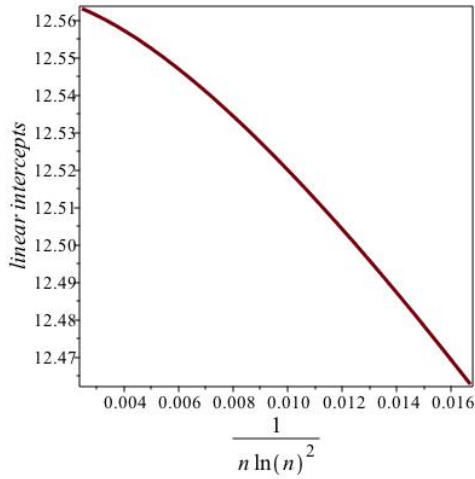


Figure 4.12: Plot of linear intercepts of ratios of  $G(t)$  vs.  $1/n \log^2 n$ .

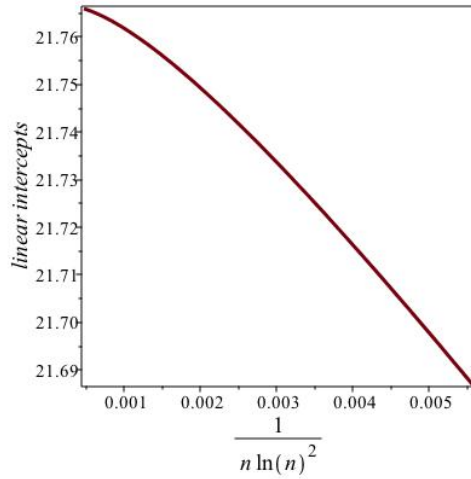


Figure 4.13: Plot of linear intercepts of ratios of  $Q(t)$  vs.  $1/n \log^2 n$ .

As we did for two stacks in series in Section 2.5.4, we now use differential

approximants to predict subsequent ratios of the series  $G(t)$  and  $Q(t)$  using the known terms. The detailed description as to how this is done is given in [124]. In this case we are extremely fortunate, in that the standard deviation of the coefficient estimates increases extremely slowly, and so we are confident in predicting 1000 extra ratios for both series which we expect to be accurate to more than 10 significant digits. That is more than enough for our purposes. Using these additional terms, we reconstruct the plot shown in Figure 4.13 in Figure 4.14, using a further 1000 ratios. Note that the locus passes through a maximum, reflecting competition between the subdominant logarithmic terms, and the linear intercepts are now decreasing with increasing  $n$ , as predicted by the asymptotic expression (4.3). In Figure 4.15, we show the same plot, but with the abscissa restricted to ratios corresponding to  $700 \leq n \leq 1100$ . The value of the ordinate at the origin in Figure 4.15 is precisely  $4\sqrt{3}\pi$ , and the extrapolated locus is convincingly going through the origin.

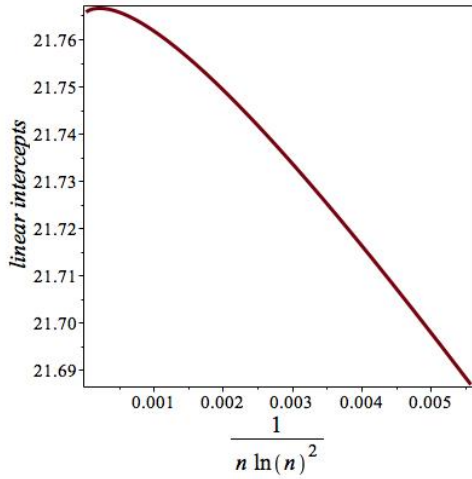


Figure 4.14: Plot of linear intercepts of ratios of  $Q(t)$  vs.  $1/n \log^2 n$ , using an extra 1000 ratios.

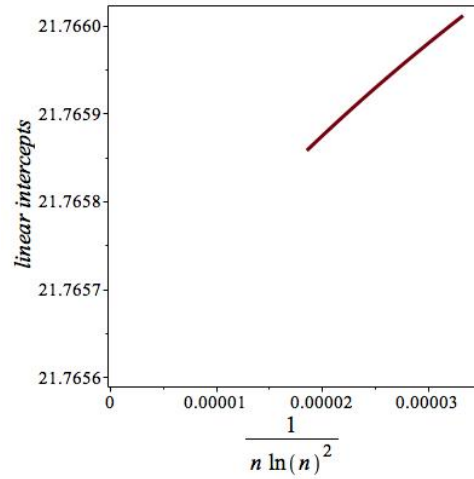


Figure 4.15: Plot of linear intercepts of ratios of  $Q(t)$  vs.  $1/n \log^2 n$ , using ratios 700 to 1100.

The corresponding plots for planar orientations, given by the generating function  $G(t)$ , are shown in figures 4.16 and 4.17, where now the value of the ordinate at the origin in Figure 4.17 is precisely  $4\pi$ , and again the extrapolated locus is convincingly going through the origin.

Now that we have good grounds to conjecture the exact value of the critical points, we are in a better position to estimate the exponent. From [106], p.385 we see that if

$$f(t) = (1 - \mu \cdot t)^{-\alpha} \left( \frac{1}{\mu \cdot t} \log \frac{1}{1 - \mu \cdot t} \right)^{\beta},$$

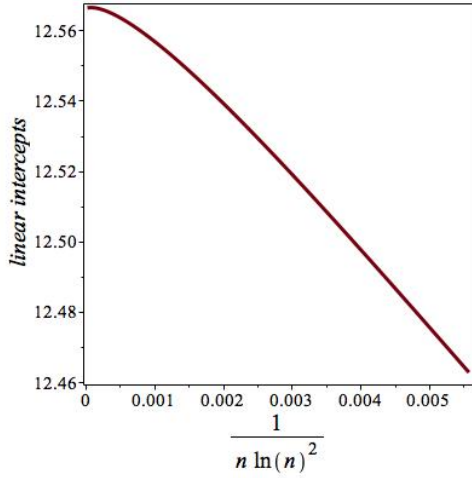


Figure 4.16: Plot of linear intercepts of ratios of  $G(t)$  vs.  $1/n \log^2 n$ , using an extra 1000 ratios.

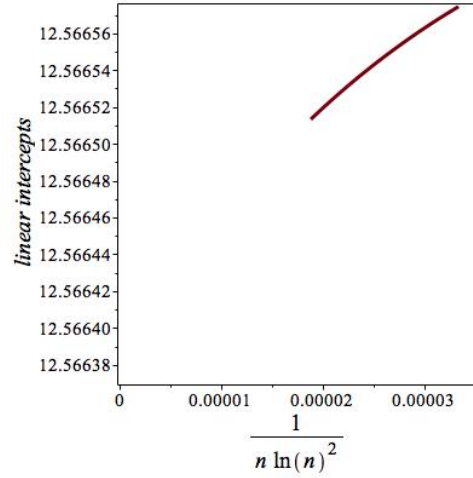


Figure 4.17: Plot of linear intercepts of ratios of  $G(t)$  vs.  $1/n \log^2 n$ , using ratios 700 to 1100.

then

$$[t^n]f(t) = \frac{\mu^n \cdot n^{\alpha-1}}{\Gamma(\alpha)} (\log n)^\beta \left( 1 + \frac{c_1}{\log n} + \frac{c_2}{\log^2 n} + \frac{c_3}{\log^3 n} + \frac{c_4}{\log^4 n} + o\left(\frac{1}{\log^4 n}\right) \right),$$

where

$$c_k = \binom{\beta}{k} \Gamma(\alpha) \frac{d^k}{ds^k} \frac{1}{\Gamma(s)} \Big|_{s=\alpha}.$$

When  $\alpha$  is a negative integer, the evaluation of the constants must be interpreted as a limiting case as the  $\Gamma$  function diverges, so that certain constants vanish. In particular, provided that  $\alpha$  is a negative integer and  $\beta$  is not zero or a positive integer, one has

$$[t^n]f(t) = \mu^n \cdot n^{\alpha-1} (\log n)^\beta \left( \frac{c_1}{\log n} + \frac{c_2}{\log^2 n} + \frac{c_3}{\log^3 n} + \frac{c_4}{\log^4 n} + o\left(\frac{1}{\log^4 n}\right) \right),$$

Then the ratio of successive coefficients is in the general case

$$r_n = \frac{[t^n]f(t)}{[t^{n-1}]f(t)} = \mu \left( 1 + \frac{\alpha-1}{n} + \frac{\beta}{n \log n} - \frac{c_1}{n \log^2 n} + o\left(\frac{1}{n \log^2 n}\right) \right),$$

but in the case that  $\alpha$  is a negative integer and  $\beta$  is not zero or a positive integer, one has

$$r_n = \frac{[t^n]f(t)}{[t^{n-1}]f(t)} = \mu \left( 1 + \frac{\alpha-1}{n} + \frac{\beta-1}{n \log n} + \frac{d}{n \log^2 n} + o\left(\frac{1}{n \log^2 n}\right) \right),$$

where  $d = 3c_2/c_1$ .

So one can estimate  $\alpha$  from the sequence

$$\alpha_n = \left( \frac{r_n}{\mu} - 1 \right) \cdot n + 1 = \alpha + \frac{\beta}{\log n} - \frac{d}{\log^2 n} + o\left(\frac{1}{\log^2 n}\right).$$

Plots of  $\alpha_n$  against  $1/\log n$  for both  $G(t)$  and  $Q(t)$  respectively are shown in figures 4.18 and 4.19, and it can be seen that having many more than 100 terms is essential. In fact the minimum in both plots occurs at around  $n = 100$ , and it is only with our extended data that the limit  $\alpha = -1$  becomes plausible.

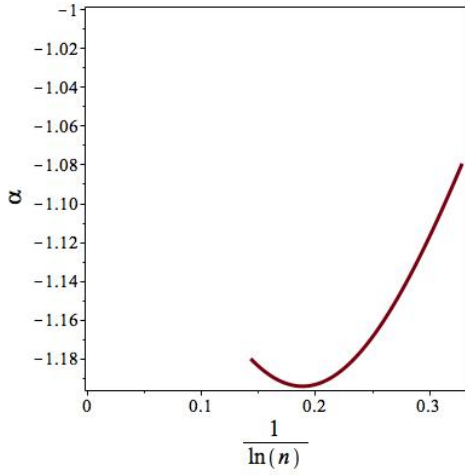


Figure 4.18: Plot of exponent  $\alpha$  estimates from  $G(t)$  vs.  $1/\log n$ , using an extra 1000 ratios.

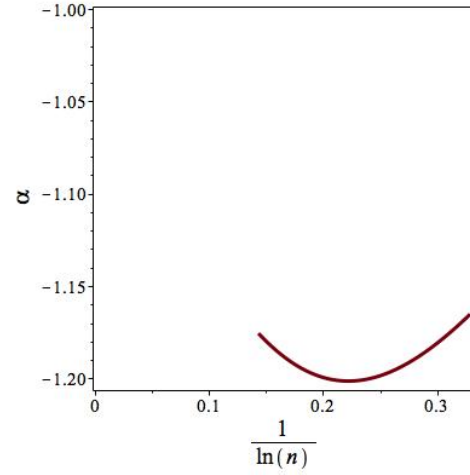


Figure 4.19: Plot of exponent  $\alpha$  estimates from  $Q(t)$  vs.  $1/\log n$ , using an extra 1000 ratios.

To take into account higher-order terms in the asymptotics, we attempted a linear fit to the assumed form (also assuming  $\alpha$  is a negative integer, otherwise  $\beta$  replaces  $\beta - 1$ ),

$$\left( \frac{r_n}{\mu} - 1 \right) \cdot n + 1 = \alpha + \frac{\beta - 1}{\log n} - \frac{d}{\log^2 n} + o\left(\frac{1}{\log^2 n}\right). \quad (4.4)$$

We did this by solving the linear system given by setting  $n = m - 1$ ,  $n = m$ ,  $n = m + 1$  in the preceding equation, and solving for  $\alpha$ ,  $\beta$ ,  $d$ , with  $m$  ranging from 20 to the maximum possible value 1100. We obtain an  $m$ -dependent sequence of estimates of the terms  $\alpha$ ,  $\beta$ ,  $d$ , which we show plotted against appropriate powers of  $1/m$ . These are shown in figures 4.20 and 4.21 for planar orientations. (The corresponding plots for 4-valent orientations are similar in appearance, so are not shown).

In this way we see that both  $\alpha$  and  $\beta$  are plausibly going to  $-1$ , as appropriate for a singularity of the form

$$\frac{c \cdot \mu \cdot t \cdot (1 - \mu \cdot t)}{\log(1 - \mu \cdot t)}.$$

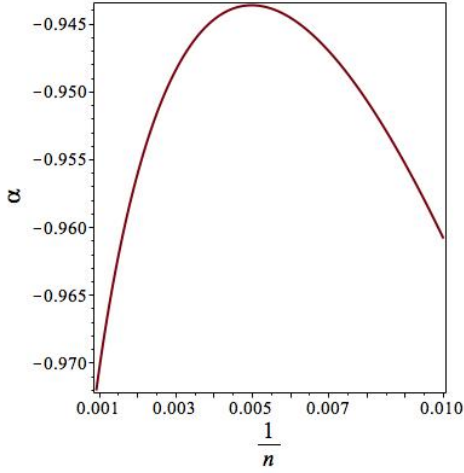


Figure 4.20: Plot of exponent  $\alpha$  estimates from eqn (4.4).

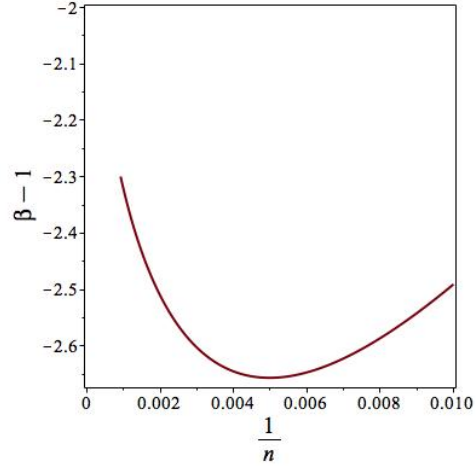


Figure 4.21: Plot of exponent  $\beta - 1$  estimates from eqn (4.4).

Finally, if we accept that  $\alpha = -1$ , we can refine the estimate of  $\beta$ , since in that case

$$\left(\frac{r_n}{\mu} - 1 + \frac{2}{n}\right) n \log n = \beta - 1 + O\left(\frac{1}{\log n}\right). \quad (4.5)$$

The result is shown in Figure 4.22 which is plausibly tending to  $\beta = -1$ , though the fact that the abscissa is  $1/\log n$  means that one would really need many more terms, around 22,000, even to get to within 0.1 on the abscissa.

## Conjecturing exact forms of $Q(t)$ and $G(t)$ .

In the last Section, we described our numerical work which resulted in the conjectures that the radii of convergence of  $Q(t)$  and  $G(t)$  are  $1/(4\sqrt{3}\pi)$  and  $1/(4\pi)$ , respectively. In other words, their coefficients have exponential growth rates  $4\sqrt{3}\pi$  and  $4\pi$ , respectively. In retrospect we learnt that for  $Q(t)$ , this was already conjectured by Kostov and Zinn-Justin in the mathematical physics literature [143][214]. Upon seeing these conjectures, Mireille



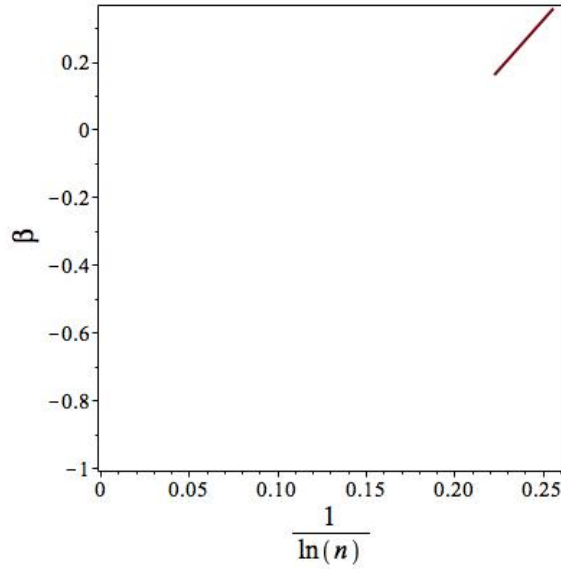


Figure 4.22: Plot of exponent  $\beta$  estimates from eqn (4.5).

Bousquet-Mélou recognised that the growth rate  $4\sqrt{3}\pi$  had appeared before in a planar map enumeration problem, namely in the series  $R(t)$  in [48], which is the unique formal power series with constant term 0 satisfying

$$t = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n} \binom{3n}{n} R^{n+1}.$$

Using this equation, we computed 100 coefficients of the series  $R(t)$ . We then searched for a linear differential equation relating the two series  $Q(t)$  and  $R(t)$  as follows: First, we computed the first 100 coefficients of each series

$$t^{k+j} \left( \frac{d}{dt} \right)^j Q(t) \quad \text{and} \quad t^{k+j} \left( \frac{d}{dt} \right)^j R(t),$$

for  $0 \leq k, j \leq 5$ . Next, we searched for a non-trivial linear combination of these series and the monomials  $1, t, \dots, t^5$  whose first 100 coefficients were equal to 0. This just involved solving 100 linear simultaneous equations. To our surprise, we found the remarkably simple *rational* relationship between the two series, namely

$$3t^2 Q(t) = t - 3t^2 - R(t).$$

As we had proved this equation up to  $O(t^{100})$ , we conjectured that it was true. For the second series,  $G(t)$ , we then searched for a similar formula.

The first step was to find a series  $R_1(t)$  defined similarly to  $R(t)$ , but with radius of convergence  $4\pi$ . Indeed, the unique series  $R_1(t)$  with constant term 0 satisfying

$$t = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n}^2 R_1^{n+1},$$

satisfies these conditions. We then searched for a linear differential equation relating  $G(t)$  and  $R_1(t)$ , and found the following *rational* relationship between the two series

$$4t^2 G(t) = t - 2t^2 - R_1(t).$$

Due to our computations, we had proved this equation up to  $O(t^{90})$ , so we conjectured that it was true.

These two conjectures are now Theorems 4.1 and 4.2, as they are proved in the following five sections.

## Simplified functional equations for quartic Eulerian orientations

In this section we will characterise the generating function  $Q(t)$  of labelled quadrangulations by a system of functional equations, which is simpler than the system of functional equations in Section 4.3.2. In particular, our new system involves only two catalytic variables  $x$  and  $y$ , as it does not include a catalytic variable counting the degree of the co-root vertex. We could not solve the old system of functional equations, however we will solve this new system in Section 4.6.

**Theorem 4.15.** *There exists a unique 3-tuple of series, denoted  $P(t, y)$ ,  $C(t, x, y)$  and  $D(t, x, y)$ , belonging respectively to  $\mathbb{Q}[[y, t]]$ ,  $\mathbb{Q}[x][[y, t]]$  and  $\mathbb{Q}[[x, y, t]]$ , and satisfying the following equations:*

$$\begin{aligned} P(t, y) &= \frac{1}{y} [x^1] C(t, x, y), \\ D(t, x, y) &= \frac{1}{1 - C\left(t, \frac{1}{1-x}, y\right)}, \\ D(t, x, y) &= 1 + y [x^{\geq 0}] \left( D(t, x, y) \left( \frac{1}{x} P\left(t, \frac{t}{x}\right) + [y^1] D(t, x, y) \right) \right), \end{aligned}$$

together with the initial condition

$$[y^1] D(t, x, y) = \frac{1}{1-x} (1 + 2t[y^2] D(t, x, y) - t([y^1] D(t, x, y))^2).$$

The generating function  $Q(t)$  that counts labelled quadrangulations by faces is

$$Q(t) = [y^1]P(t, y) - 1.$$

By Lemma 4.4, the series  $Q(t)$  also counts quartic Eulerian orientations by vertices.

### Remark

With the conditions on the series  $P$ ,  $C$  and  $D$ , the operations that occur in the above equations are always well defined:

- the coefficient of  $x$  in  $C(t, x, y)$  lies in  $\mathbb{Q}[[y, t]]$ ,
- the series  $C(t, 1/(1-x), y)$  indeed lies in  $\mathbb{Q}[[x, y, t]]$  (upon expanding the powers of  $1/(1-x)$  as series in  $x$ ),
- denoting by  $p_{j,n} \in \mathbb{Q}$  the coefficient of  $y^j t^n$  in  $P(t, y)$ , and by  $d_{j,n}(x) \in \mathbb{Q}[[x]]$  the coefficient of  $y^j t^n$  in  $D$ , the quantity

$$\begin{aligned} \frac{1}{x} D(t, x, y) P\left(t, \frac{t}{x}\right) &= \left( \sum_{j,n \geq 0} d_{j,n}(x) y^j t^n \right) \left( \sum_{i,m \geq 0} p_{i,m} \frac{1}{x^{i+1}} t^{i+m} \right) \\ &= \sum_{j,N \geq 0} y^j t^N \sum_{i+m+n=N} p_{i,m} d_{j,n}(x) \frac{1}{x^{i+1}} \end{aligned}$$

is a series in  $y$  and  $t$  whose coefficients are Laurent series in  $x$  (because  $i, m$  and  $n$  are bounded). It thus makes sense to extract its non-negative part in  $x$ , which will lie in  $\mathbb{Q}[[x, y, t]]$ .

The series  $P$  and  $D$  of Theorem 4.15 count patches and D-patches, as described in Definition 4.7.  $C$  counts C-patches, which we now define. We also include the definitions of patches and D-patches. See Figure 4.23 for an illustration of these three definitions.

**Definition 4.16.** A *patch* is a labelled map in which each inner face has degree 4, and the vertices around the outer face are alternately labelled 0 and 1.

A *C-patch* is a patch satisfying two additional conditions: all neighbours of the root vertex are labelled 1, and the root corner is the only outer corner at the root vertex. By convention, the atomic patch is *not* a C-patch.

*D-patches* resemble patches but may include digons. More precisely, a D-patch is a labelled map in which each inner face has degree 2 or 4, those of degree 2 being incident to the root vertex, and the vertices around the outer face are alternately labelled 0 and 1. We also require that all neighbours of the root vertex are labelled 1.

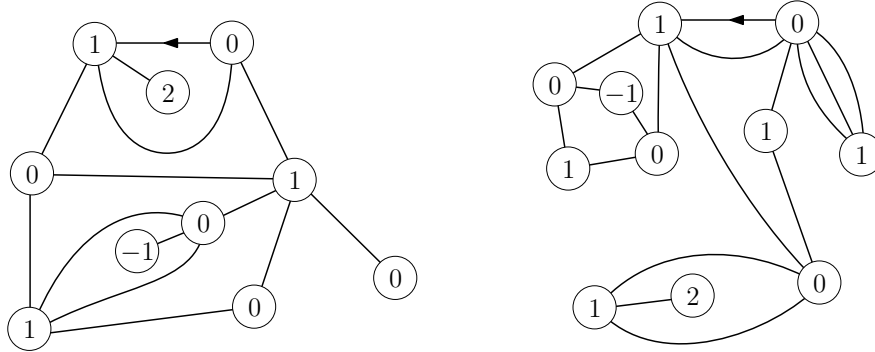


Figure 4.23: Left: a C-patch (and hence, also a patch) which contributes  $t^5y^4$  to the generating function  $P(t, y)$  and  $t^5x^3y^4$  to the generating function  $C(t, x, y)$ . Right: a D-patch which contributes  $t^4x^3y^6$  to the generating function  $D(t, x, y)$ .

We define  $P(t, y)$ ,  $C(t, x, y)$  and  $D(t, x, y)$  to be respectively the generating functions of patches, C-patches and D-patches, where  $t$  counts inner quadrangles,  $y$  the outer degree (halved), and  $x$  either the degree of the root vertex (for C-patches) or the number of inner digons (for D-patches). Comparing with the functions in Section 4.3.2, we see that one parameter, namely the degree of the co-root vertex, is no longer involved here. The series  $P$ ,  $C$  and  $D$  actually belong to the rings described by Theorem 4.15:

- for  $P$  it suffices to observe that there are finitely many patches with  $n$  inner quadrangles and outer degree  $2j$ ,
- for  $D$  we observe that there are finitely many D-patches with  $n$  inner quadrangles,  $i$  inner digons and outer degree  $2j$ ,
- finally for  $C$ , we note that a C-patch with  $n$  inner quadrangles cannot have a root vertex of degree larger than  $1 + 4n$ , because all non-root corners at the root vertex must belong to an inner quadrangle (by the second condition of Definition 4.16). This explains the polynomiality of  $[t^n]C$  in  $x$  (and yields in fact a smaller ring than  $\mathbb{Q}[x][[y, t]]$ , namely  $(\mathbb{Q}[[y]][x])[t]$ , but this won't be needed).

In the next 5 lemmas, we prove that the series that we have defined satisfy the 5 equations of Theorem 4.15. We will finish the section by proving uniqueness of  $C$ ,  $P$  and  $D$ .

**Lemma 4.17.** *The generating functions  $P(t, y)$  and  $C(t, x, y)$  satisfy the equation*

$$P(t, y) = \frac{1}{y}[x^1]C(t, x, y).$$

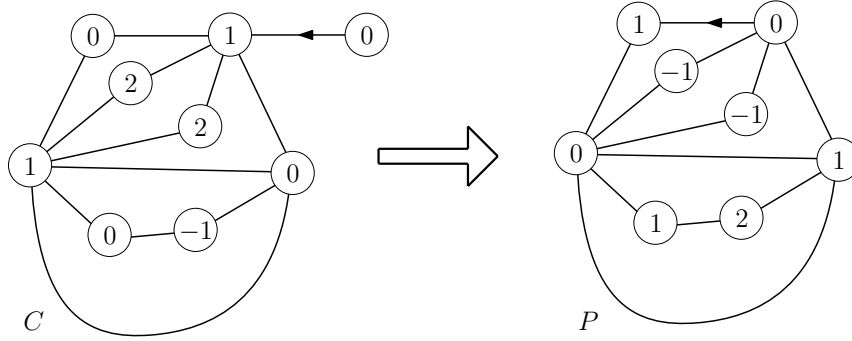


Figure 4.24: The transformation of  $C$  into  $P$  used in the proof of Lemma 4.17.

*Proof.* Let  $C$  be any C-patch counted by  $[x^1]\mathcal{C}(t, x, y)$ , that is, in which the root vertex has degree 1. We construct a new patch  $P$  from  $C$ , as illustrated in Figure 4.24: we delete the root edge and root vertex of  $C$ , replace each label  $\ell$  with  $1 - \ell$ , and finally root  $P$  at the outer edge of  $C$  following the root edge of  $C$  anticlockwise. Then the new labelled map  $P$  is indeed a patch. If  $C$  contains only one edge then  $P$  is the atomic map. The outer degree has decreased by 2, while the number of inner quadrangles is unchanged. Finally, the transformation from  $C$  to  $P$  is reversible. This proves the lemma.  $\square$

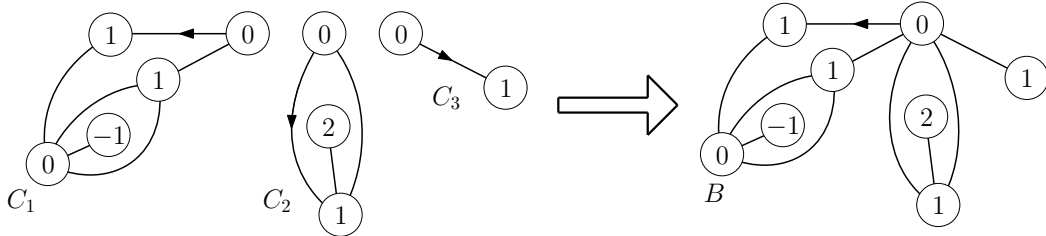


Figure 4.25: A sequence of C-patches gives rise to a B-patch, as in Lemma 4.18.

**Lemma 4.18.** *The generating functions  $\mathcal{D}(t, x, y)$  and  $\mathcal{C}(t, x, y)$  satisfy the equation*

$$\mathcal{D}(t, x, y) = \frac{1}{1 - \mathcal{C}\left(t, \frac{1}{1-x}, y\right)}.$$

*Proof.* Recall that C-patches satisfy two conditions: all neighbours of the root vertex have label 1, and the root vertex is only incident once to the root face. By attaching a sequence of C-patches at their root vertex, as shown in Figure 4.25, we form a *B-patch*, that is, a patch satisfying only the first of these conditions. The associated generating function is

$$\mathcal{B}(t, x, y) = \frac{1}{1 - \mathcal{C}(t, x, y)}.$$

As before,  $t$  counts inner quadrangles,  $x$  the degree of the root vertex and  $y$  the outer degree (halved).

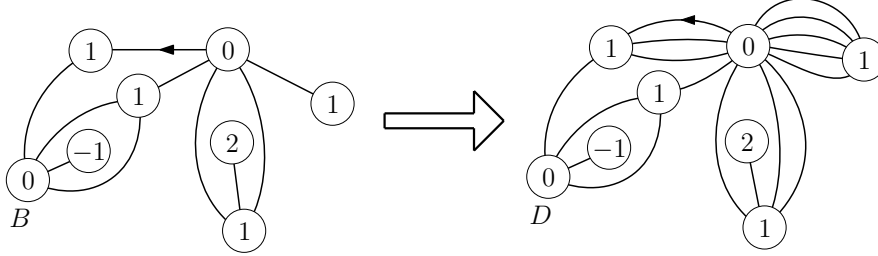


Figure 4.26: The transformation from a B-patch to a D-patch, as in Lemma 4.18.

Now in order to construct a D-patch, it suffices to take a B-patch and inflate every edge which is incident to the root vertex into a sequence of digons, as shown in Figure 4.26. This explains the transformation  $x \mapsto 1/(1-x)$  occurring in the lemma. In this way the variable  $x$  now counts digons of D-patches.  $\square$

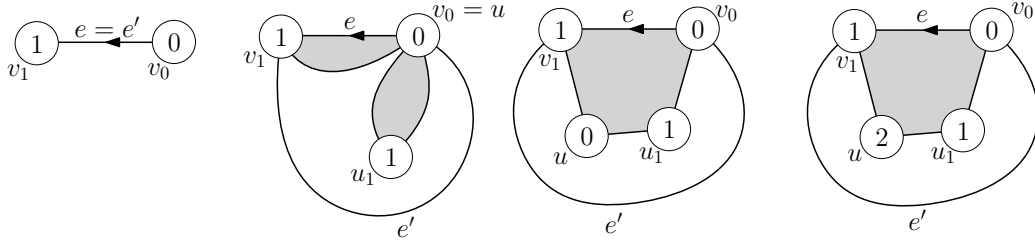


Figure 4.27: The four different types of patches which contribute to the generating function  $[y^1]\mathcal{C}(t, x, y)$ . In the third and fourth cases it is possible that  $u_1 = v_1$ . The shaded area represents a labelled quadrangulation.

**Lemma 4.19.** *The generating function  $\mathcal{D}(t, x, y)$  satisfies the equation*

$$[y^1]\mathcal{D}(t, x, y) = \frac{1}{1-x} (1 + 2t[y^2]\mathcal{D}(t, x, y) - t([y^1]\mathcal{D}(t, x, y))^2).$$

*Proof.* We will show that

$$[y^1]\mathcal{C}(t, x, y) = x (1 + 2t[y^2]\mathcal{C}(t, x, y) + t([y^1]\mathcal{C}(t, x, y))^2), \quad (4.6)$$

from which the desired result follows using Lemma 4.18, while observing that  $\mathcal{C}(t, x, 0) = 0$ . Let  $C$  be any C-patch counted by  $[y^1]\mathcal{C}(t, x, y)$ , that is, having outer degree 2. Let  $e$  be the root edge of  $C$ , let  $e'$  be the other outer edge of

$C$  and let  $v_0$  and  $v_1$  be the root vertex and co-root vertex respectively. We consider four cases, illustrated in Figure 4.27.

In the first case  $e = e'$ . Since the outer degree of  $C$  is 2, this is only possible if  $e$  is the only edge in  $C$ , so this case simply contributes  $x$  to  $[y^1]\mathcal{C}(t, x, y)$ . For the other three cases, let  $Q$  be the map remaining when  $e'$  is removed (this is the shaded area in Figure 4.27). Then the outer degree of  $Q$  must be 4, so that  $Q$  is a quadrangulation. Let the vertices around the outer face of  $Q$  be  $v_0, v_1, u$  and  $u_1$  in anticlockwise order. Note that  $v_1$  and  $u_1$  must both be labelled 1 since they are adjacent to  $v_0$  and  $C$  is a C-patch.

The second case we consider is when  $u = v_0$ . Then  $Q$  can be separated into two C-patches with outer degree 2, hence this case contributes

$$xt([y^1]\mathcal{C}(t, x, y))^2$$

to  $[y^1]\mathcal{C}(t, x, y)$ . The factor  $xt$  appears because the number of inner quadrangles in  $Q$  and the degree of the root vertex of  $Q$  are each one less than the equivalent numbers in  $C$ .

The third case is when  $u \neq v_0$ , but  $u$  is labelled 0. Then  $Q$  can be any C-patch with outer degree 4. Hence this case contributes

$$xt[y^2]\mathcal{C}(t, x, y).$$

In the fourth and final case,  $u$  is labelled 2, and therefore it cannot be equal to  $v_0$ . In this case  $Q$  is not a patch because of this label 2 on its outer face. But we construct a new map  $Q'$  from  $Q$  by replacing every label  $\ell$  in  $Q$  with  $2 - \ell$ , except for the label at the root vertex, which remains 0. Then  $Q'$  is still a labelled map, all neighbours of the root vertex are still labelled 1, and the root face is only incident once to the root vertex. Hence,  $Q'$  can be any C-patch with outer degree 4, so this case contributes

$$xt[y^2]\mathcal{C}(t, x, y).$$

Adding the contributions from the four cases yields (4.6), which, in turn, yields the desired result using Lemma 4.18.  $\square$

Before we can prove that the next equation holds (the one that expresses  $\mathcal{D}(t, x, y)$ ), we need to introduce minus-patches, subpatches and a contraction operation. This contraction operation was already used in [91].

**Definition 4.20.** A *minus-patch* is a map obtained from a patch by replacing each label  $\ell$  with  $-\ell$ .

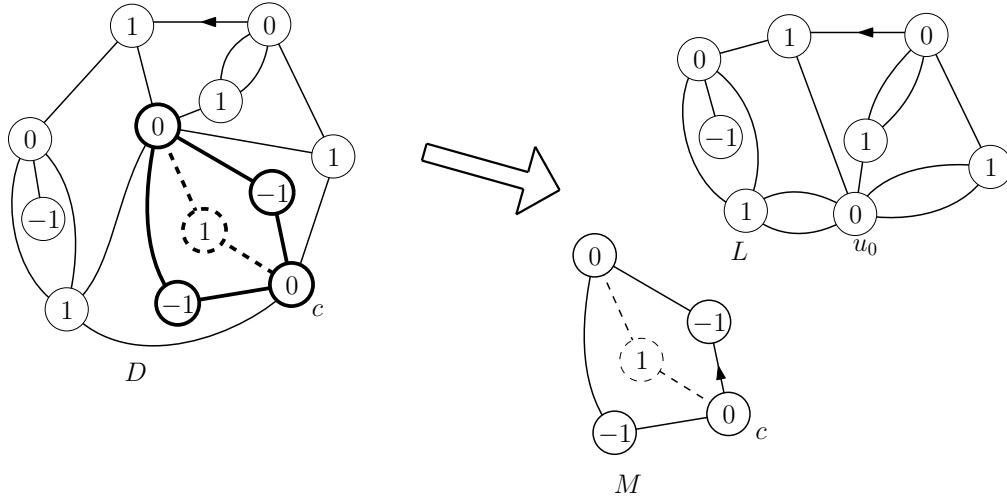


Figure 4.28: Left: a D-patch with a chosen outer corner  $c$  labelled 0. The minus-subpatch  $M$  of  $D$  rooted at  $c$  is highlighted in  $D$  and shown separately in the middle. The submap  $M'$  is obtained from  $M$  by deleting the dashed vertex and edges. Right: the labelled map  $L$  constructed from  $D$  by contracting  $M$  to a single vertex  $u_0$ .

Clearly these are equinumerous with patches.

We now describe a way to extract a minus-subpatch from a D-patch. This is identical to Definition 4.5, in the case where the labelled map  $L$  is a D-patch. This definition is illustrated in Figure 4.28.

**Definition 4.21.** Let  $D$  be a D-patch and let  $c$  be an outer corner of  $D$  at a non-root vertex  $v$  labelled 0. We define the *minus-subpatch of  $D$  rooted at  $c$*  as follows. First, let  $M'$  be the maximal submap of  $D$  that contains  $v$  and consists of vertices labelled 0 or less. Let  $M$  be the submap of  $D$  that contains  $M'$  and all edges and vertices within its boundary (assuming the root face is drawn as the infinite face). The map  $M$ , which we root at the corner inherited from  $c$ , is the *minus-subpatch of  $D$  rooted at  $c$* .

It follows immediately from this definition that  $M$  and  $M'$  share the same outer face. Moreover, all inner faces of  $M$  are also inner faces of  $D$ . Since the root vertex of  $D$  is only adjacent to vertices labelled 1, it cannot be a vertex of  $M'$ , so it cannot be a vertex of  $M$  either. Hence all inner faces of  $M$  are quadrangles, since all digons in  $D$  are incident to its root. All outer vertices of  $M$  must also be outer vertices of  $M'$ , so they have non-positive labels. Let us prove that these labels can only be 0 and  $-1$ . For any outer vertex  $u$  of  $M$ , there is some face  $F$  of  $D$ , containing  $u$ , which is not a face of  $M$ . If  $F$  is the outer face of  $D$ , with labels 0 and 1, then the label of  $u$ , being non-



positive, can only be 0. The face  $F$  cannot be a digon, otherwise  $u$  would be the vertex labelled 0, and thus would be the root vertex of  $D$ , while we have shown that this vertex is not in  $M$ . Finally, if  $F$  is an inner quadrangle of  $D$ , then it must contain a vertex  $u'$  with label at least 1 (otherwise  $F$  would be contained in  $M$ ). Since  $u$  and  $u'$  are incident to the same quadrangle  $F$ , and  $u$  has a non-positive label, this label can only be  $-1$  or  $0$ . Hence the outer vertices of  $M$  are all labelled  $0$  or  $-1$ , so  $M$  is a minus-patch, in the sense of Definition 4.20. This justifies the terminology *minus-subpatch* of  $D$ .

Now, every edge in  $D$  which connects a vertex in  $M$  to a vertex not in  $M$  must have endpoints labelled  $0$  (in  $M$ ) and  $1$  (not in  $M$ ). Conversely, from every corner  $c$  labelled  $0$  on the outer face of  $M$  there must start an edge ending at a vertex labelled  $1$ . Otherwise, the face of  $D$  that contains  $c$  would contain a vertex labelled  $-1$  and would have degree larger than  $4$ , which is impossible. We can thus form a new labelled map  $L$  by contracting all of  $M$  to a single vertex  $u_0$  labelled  $0$  (Figure 4.28). This vertex  $u_0$  is only adjacent to vertices labelled  $1$  in  $L$ , and the number of inner digons containing  $u_0$  is half the outer degree of  $M$ . Note that the contracted map  $L$  is not a D-patch (unless the minus-patch  $M$  is only a single vertex), because there are digons incident to the non-root vertex  $u_0$ . Nevertheless, we will use this contraction in the proof of the next lemma, by transforming  $L$  further under certain additional assumptions.

For the moment, let us show that the transformation that maps the D-patch  $D$ , with its marked outer corner  $c$ , to the pair  $(M, L)$ , is reversible, provided the outer corner of  $L$  inherited from  $c$  (and still denoted  $c$ ) is marked. If  $M$  is atomic, there is nothing to do. Otherwise, let  $u_0$  be the vertex located at the corner  $c$  in  $L$ . Let  $e_1, e_2, \dots, e_d$  be the edges attached to  $u_0$ , in anticlockwise order around  $u_0$ , starting from the corner  $c$  (Figure 4.29). If  $M$  has outer degree  $2j$ , then  $u_0$  is incident to  $j$  inner digons. We now erase the vertex  $u_0$  from  $L$ , so that the half-edges  $e_i$  are dangling. We connect them to the outer corners of  $M$  labelled  $0$  in the following way: we first attach  $e_1$  to the root corner of  $M$ , and then proceed anticlockwise around  $M$ , connecting  $e_{i+1}$  to the next corner of  $M$  labelled  $0$  if the corner of  $L$  defined by  $e_i$  and  $e_{i+1}$  belongs to an inner digon of  $L$ , and to the same corner as  $e_i$  otherwise. In this way we only create quadrangles.

We are now ready to prove the most complex of our equations.

**Lemma 4.22.** *The generating functions  $P(t, y)$  and  $D(t, x, y)$  satisfy the equation*

$$D(t, x, y) = 1 + y [x^{\geq 0}] \left( D(t, x, y) \left( \frac{1}{x} P \left( t, \frac{t}{x} \right) + [y^1] D(t, x, y) \right) \right).$$

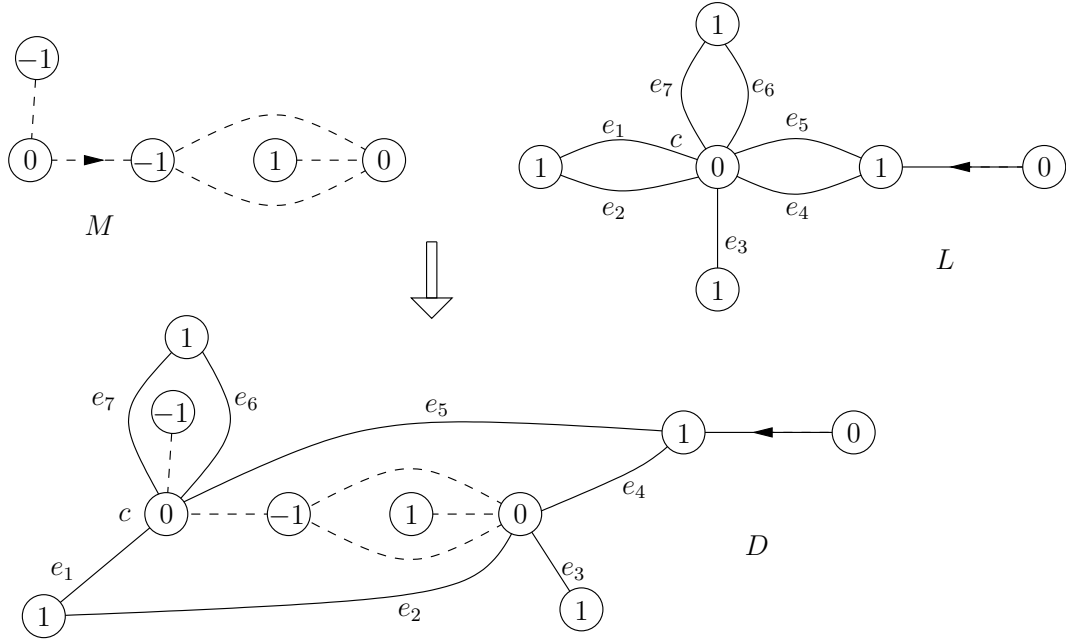


Figure 4.29: How to reconstruct the marked D-patch  $(D, c)$  from the minus-patch  $M$  (dashed edges) and  $(L, c)$ . Here,  $M$  has outer degree  $2j = 6$ , and the vertex at the corner  $c$  has degree  $d = 7$ .

*Proof.* Let  $D$  be any D-patch. We will consider three cases, illustrated in Figure 4.30. In the first case,  $D$  is atomic, and contributes 1 to  $D(t, x, y)$ . For the other cases, let  $v_0$  be the root vertex, let  $c_1$  and  $c_0$  be the next two corners clockwise around the outer face from the root corner, and let  $u_0$  be the vertex associated with  $c_0$ . The second case that we consider is when  $u_0 = v_0$ . In this case we can split the D-patch into two D-patches at  $v_0$ . One of them has outer degree 2 and the other can be any D-patch. Hence, in this case we get the contribution

$$y([y^1]D(t, x, y))D(t, x, y).$$

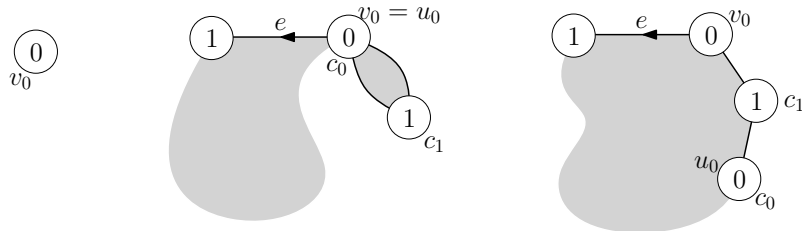


Figure 4.30: The three different types of D-patches. In the third case it is possible that the two displayed vertices labelled 1 are the same vertex.

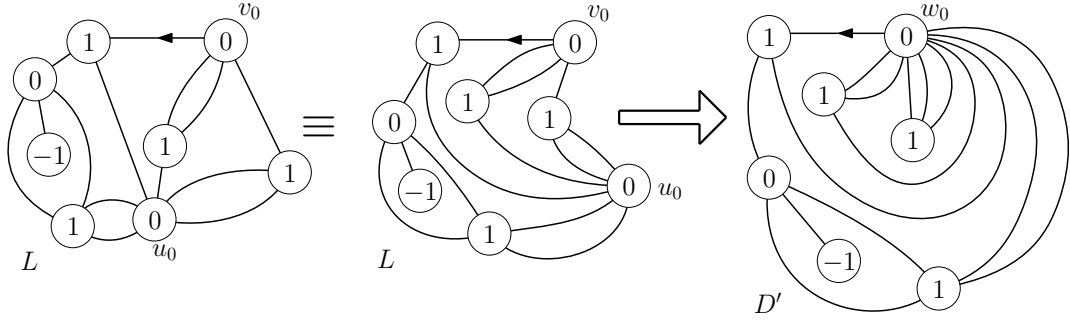


Figure 4.31: The transformation of  $L$  into a D-patch  $D'$  used in the proof of Lemma 4.22.

In the third and final case,  $u_0 \neq v_0$ . Let  $M$  be the minus-subpatch of  $D$  rooted at  $c_0$  and let  $L$  be the map obtained by contracting  $M$  to a single vertex  $u_0$  labelled 0, as described above the lemma (Figure 4.28). We note that  $D$  and  $L$  have the same outer degree since no edge of the boundary of  $D$  has been contracted. We have described above the lemma how to recover  $D$  from  $M$  and  $L$  (note that the choice of the corner  $c = c_0$  is canonical here). In general,  $L$  is not a D-patch, as there may be inner digons incident to the non-root vertex  $u_0$ . But we can construct a D-patch  $D'$  from  $L$  by moving  $u_0$  towards  $v_0$  in the outer face until these two vertices merge into a new root vertex  $w_0$  (Figure 4.31; we do not merge any edges). This creates an extra digon at  $w_0$ , in addition to those that were incident to  $u_0$  and  $v_0$ . Now all digons in  $D'$  are incident to the root vertex  $w_0$ , and all neighbours of  $w_0$  have label 1. Hence  $D'$  is a D-patch.

Note that we can reconstruct  $L$  from  $D'$ , provided we know the outer degree of  $M$ , that is, the number of inner digons that will be incident to  $u_0$  once  $w_0$  is split into two vertices  $v_0$  and  $u_0$ . Hence the transformation that sends a D-patch  $D$  of the third type to a pair  $(M, D')$ , such that the number of inner digons of  $D'$  is larger than half the outer degree of  $M$ , is bijective. The parameters of interest behave as follows:

- the number of inner quadrangles of  $D$  is the sum of the corresponding numbers in  $M$  and  $D'$ , plus half the outer degree of  $M$ ,
- the number of inner digons in  $D$  is the number of inner digons in  $D'$ , minus half the outer degree of  $M$ , minus 1,
- the outer degree of  $D$  is the outer degree of  $D'$  plus 2.

Hence, the contribution from the third case is

$$\frac{y}{x} [x^{>0}] \left( D(t, x, y) P \left( t, \frac{t}{x} \right) \right) = y [x^{\geq 0}] \left( \frac{1}{x} D(t, x, y) P \left( t, \frac{t}{x} \right) \right).$$

Adding the contributions from the three cases together yields the desired result.  $\square$

**Lemma 4.23.** *The generating function  $Q(t)$  is given by*

$$Q(t) = [y^1]P(t, y) - 1.$$

*Proof.* Let  $Q$  be any labelled quadrangulation. The outer face may contain a label  $-1$  or  $2$ , hence  $Q$  is not necessarily a patch. Let  $P$  be the map constructed from  $Q$  by adding an edge  $e'$  between the root vertex and co-root vertex in the outer face of  $Q$ , so that  $e'$  and the root edge  $e$  are the only outer edges of  $P$ . Then  $P$  can be any patch with outer degree 2, except for the patch with only one edge. Hence the possible patches  $P$  are counted by  $([y^1]P(t, y) - 1)$ . Since the number of inner faces of  $P$  is equal to the total number of faces of  $Q$ , this expression is exactly equal to  $Q(t)$ . This concludes the proof.  $\square$

*Proof of Theorem 4.15.* We have now proved the five functional equations. It remains to prove that, together with the conditions on the rings that contain  $P$ ,  $C$  and  $D$ , they determine these three series. Let us denote by  $p_{j,n}$  the coefficient of  $y^j t^n$  in  $P(t, y)$ , and similarly for  $C$  and  $D$ . These quantities should be thought of respectively as elements of  $\mathbb{Q}$  (for  $P$ ), of  $\mathbb{Q}[x]$  (for  $C$ ) and of  $\mathbb{Q}[[x]]$  (for  $D$ ). We will prove by induction on  $N \geq 0$  that

- $p_{j,n}$  is completely determined for  $j + n < N$ ,
- $c_{j,n}$  and  $d_{j,n}$  are completely determined for  $j + n \leq N$ .

When  $N = 0$ , there is nothing to prove for  $P$ . The third equation of the system shows that  $D - 1$  is a multiple of  $y$ . That is, not only  $d_{0,0} = 1$ , but in fact we also know that  $d_{0,n} = 0$  for  $n \geq 1$ . The second equation then tells us that  $C$  is a multiple of  $y$ , so that  $c_{0,n} = 0$  for  $n \geq 0$ . Now assume that the induction hypothesis holds for some  $N \geq 0$ , and let us prove it for  $N + 1$ .

We begin with the series  $D$ . Of course it suffices to determine the coefficients  $d_{j,n}$  for  $j + n = N + 1$ . We have already explained that  $d_{0,N+1} = 0$ , so we take  $j \geq 1$ . The third equation of the system expresses  $d_{j,N+1-j}$  in terms of the series  $d_{j-1,m}$  (for  $m \leq N + 1 - j$ ),  $p_{k,\ell}$  (for  $k + \ell \leq N + 1 - j$ ) and  $d_{1,m}$  (for  $m \leq N + 1 - j$ ). If  $j \geq 2$ , these series are known, by the induction hypothesis, and thus  $d_{j,N+1-j}$  is completely determined. As argued below Theorem 4.15, it belongs to  $\mathbb{Q}[[x]]$ . To determine the final coefficient  $d_{1,N}$ , we resort to the fourth equation, which expresses  $d_{1,N}$  in terms of  $d_{2,N-1}$  (which we have just determined) and the series  $d_{1,m}$  for  $m \leq N - 1$  (which are known by the induction hypothesis). Again,  $d_{1,N}$  belongs to  $\mathbb{Q}[[x]]$ .

Hence for  $j + n \leq N + 1$ , the coefficients  $d_{j,n}$  are uniquely determined and hence must count D-patches with outer degree  $2j$  and  $n$  quadrangles. Since we know that the generating functions of C-patches and D-patches are related by the second equation (see Lemma 4.18), this forces the coefficients  $c_{j,n}$ , for  $j + n \leq N + 1$ , to count C-patches. Hence they are also fully determined (and are *polynomials* in  $x$ ). Finally, the first equation of the system shows that the numbers  $p_{j,n}$  are also determined for  $j + n \leq N$  (we cannot go up to  $N + 1$  because of the division by  $y$ ).

This concludes our induction.  $\square$

## Solution for quartic Eulerian orientations

We are now about to solve the system of Theorem 4.15, thus proving, in particular, that the generating function  $Q(t)$  of quartic Eulerian orientations is indeed given by Theorem 4.1. The third equation of the system suggests that we should consider the series  $P(t, ty)$  rather than  $P(t, y)$ . In turn, this leads us to apply the same transformation to the series  $C$  and  $D$ . More precisely, let us consider

$$\mathcal{P}(t, y) = t P(t, ty), \quad \mathcal{C}(t, x, y) = C(t, x, ty), \quad \mathcal{D}(t, x, y) = D(t, x, ty). \quad (4.7)$$

Of course, if we determine  $\mathcal{P}$ ,  $\mathcal{C}$  and  $\mathcal{D}$ , then  $P$ ,  $C$  and  $D$  are completely determined as well.

The solution below has been *guessed*, and then of course checked. The first step was the discovery of the connection between the generating function  $Q$  and the series  $R$  coming from [48]. To our knowledge, this is the first time that series like  $C$  or  $D$  appear in combinatorial enumeration.

**Theorem 4.24.** *Let  $R(t) \equiv R$  be the unique formal power series with constant term 0 satisfying*

$$t = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n} \binom{3n}{n} R^{n+1}. \quad (4.8)$$

*Then the above series  $\mathcal{P}$ ,  $\mathcal{C}$  and  $\mathcal{D}$  are:*

$$\mathcal{P}(t, y) = \sum_{n \geq 0} \sum_{j=0}^n \frac{1}{n+1} \binom{2n-j}{n} \binom{3n-j}{n} y^j R^{n+1}, \quad (4.9)$$

$$\mathcal{C}(t, x, y) = 1 - \exp \left( - \sum_{n \geq 0} \sum_{j=0}^n \sum_{i=0}^{2n-j} \frac{1}{n+1} \binom{2n-j}{n} \binom{3n-i-j}{n} x^{i+1} y^{j+1} R^{n+1} \right),$$

$$\mathcal{D}(t, x, y) = \exp \left( \sum_{n \geq 0} \sum_{j=0}^n \sum_{i \geq 0} \frac{1}{n+1} \binom{2n-j}{n} \binom{3n+i-j+1}{2n-j} x^i y^{j+1} \mathbf{R}^{n+1} \right). \quad (4.10)$$

The generating function of quartic Eulerian orientations, counted by vertices, is

$$\mathbf{Q}(t) = \frac{1}{3t^2} (t - 3t^2 - \mathbf{R}(t)).$$

*Proof.* We take for  $\mathcal{P}$ ,  $\mathcal{C}$  and  $\mathcal{D}$  the above series, and define  $\mathbf{P}$ ,  $\mathbf{C}$  and  $\mathbf{D}$  by (4.7). Since  $\mathbf{R} = O(t)$ , these three series are easily seen to belong respectively to the rings  $\mathbb{Q}[[y, t]]$ ,  $\mathbb{Q}[x][[y, t]]$  and  $\mathbb{Q}[[x, y, t]]$ , as required by Theorem 4.15. Thus it suffices to check that the first four equations of Theorem 4.15 hold, or, equivalently, that

$$\begin{aligned} \mathcal{P}(t, y) &= \frac{1}{y} [x^1] \mathcal{C}(t, x, y), \\ \mathcal{D}(t, x, y) &= \frac{1}{1 - \mathcal{C}\left(t, \frac{1}{1-x}, y\right)}, \\ \mathcal{D}(t, x, y) &= 1 + y [x^{\geq 0}] \left( \mathcal{D}(t, x, y) \left( \frac{1}{x} \mathcal{P}\left(t, \frac{1}{x}\right) + [y^1] \mathcal{D}(t, x, y) \right) \right), \\ [y^1] \mathcal{D}(t, x, y) &= \frac{1}{1-x} (t + 2[y^2] \mathcal{D}(t, x, y) - ([y^1] \mathcal{D}(t, x, y))^2). \end{aligned} \quad (4.11)$$

Note that the first three equations do not involve explicitly the variable  $t$ : we will prove them without resorting to the definition (4.8) of  $\mathbf{R}$ .

The first equation is straightforward. For the second one, it suffices to prove that for all  $j \leq n$ ,

$$\sum_{i=0}^{2n-j} \binom{3n-i-j}{n} \frac{1}{(1-x)^{i+1}} = \sum_{i \geq 0} \binom{3n+i-j+1}{2n-j} x^i.$$

This follows by expanding the left-hand side in  $x$  and using the classical identity:

$$\sum_{i=0}^{k-n} \binom{k-i}{n} \binom{\ell+i}{\ell} = \binom{k+\ell+1}{n+\ell+1}. \quad (4.12)$$

We now come to the third, and most interesting, equation. Our first observation is that, in the expression (4.10) of  $\mathcal{D}(t, x, y)$ , the sum over  $i$  is a

rational function of  $x$ :

$$\begin{aligned} \sum_{i \geq 0} \binom{3n+i-j+1}{2n-j} x^i &= \sum_{k \geq n+1} \binom{2n-j+k}{2n-j} x^{k-n-1} \\ &= \frac{1}{x^{n+1}(1-x)^{2n-j+1}} - \sum_{\ell=0}^n \binom{3n-\ell-j}{2n-j} \frac{1}{x^{\ell+1}}. \end{aligned} \quad (4.13)$$

We note that the sum over  $\ell$  in the above expression is a polynomial in  $1/x$ , with no constant term. Let us denote it by  $L_{j,n}(1/x)$ . The expression of  $\mathcal{D}$  thus reads

$$\begin{aligned} \mathcal{D}(t, x, y) &= \exp \left( \sum_{n \geq 0} \sum_{j=0}^n \frac{1}{n+1} \binom{2n-j}{n} y^{j+1} \mathbf{R}^{n+1} \left( \frac{1}{x^{n+1}(1-x)^{2n-j+1}} - L_{j,n}(1/x) \right) \right) \\ &= \exp (A(\mathbf{U}, z) - B(\mathbf{R}, 1/x, y)), \end{aligned} \quad (4.14)$$

where

$$\begin{aligned} \mathbf{U} &= \frac{\mathbf{R}}{x(1-x)^2}, \quad z = (1-x)y, \\ A(u, z) &= \sum_{n \geq 0} \sum_{j=0}^n \frac{1}{n+1} \binom{2n-j}{n} z^{j+1} u^{n+1} \end{aligned} \quad (4.15)$$

and

$$B(r, 1/x, y) = \sum_{n \geq 0} \sum_{j=0}^n \frac{1}{n+1} \binom{2n-j}{n} L_{j,n}(1/x) y^{j+1} r^{n+1}.$$

By extracting the coefficient of  $y$  from (4.14), we find

$$\begin{aligned} [y^1] \mathcal{D}(t, x, y) &= (1-x) \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n} \mathbf{U}^{n+1} - \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n} L_{0,n}(1/x) \mathbf{R}^{n+1}, \\ &= (1-x) \mathbf{U} \text{Cat}(\mathbf{U}) - \frac{1}{x} \mathcal{P} \left( t, \frac{1}{x} \right), \end{aligned} \quad (4.16)$$

where  $\text{Cat}(u)$  is the Catalan series  $\sum \binom{2n}{n} \frac{u^n}{n+1}$  and  $\mathcal{P}$  is given by (4.9) (we have used the fact that  $\binom{2n}{n} \binom{3n-\ell}{2n} = \binom{2n-\ell}{n} \binom{3n-\ell}{n}$ ). The identity (4.11) that we have to prove thus reads

$$1 = [x^{\geq 0}] (\mathcal{D}(t, x, y) (1 - z \mathbf{U} \text{Cat}(\mathbf{U}))),$$

where we still denote  $z = (1-x)y$ . Equivalently, in view of (4.14):

$$1 = [x^{\geq 0}] \left( \exp(A(\mathbf{U}, z)) (1 - z \mathbf{U} \text{Cat}(\mathbf{U})) \exp(-B(\mathbf{R}, 1/x, y)) \right).$$

We will prove below in Lemma 4.25 that

$$\exp(A(U, z)) (1 - zU \text{Cat}(U)) = 1,$$

which, given that  $B(R, 1/x, y)$  only involves negative powers of  $x$ , concludes the proof of the third identity.

Consider now the fourth equation of the system. Given that the second equation holds, what we need to prove can be rewritten as:

$$[y^1]\mathcal{C}(t, x, y) = x \left( t + 2[y^2]\mathcal{C}(t, x, y) + ([y^1]\mathcal{C}(t, x, y))^2 \right).$$

Let us write  $\mathcal{C}(t, x, y) = 1 - \exp(-T(t, x, y))$ . Then the above identity reads:

$$[y^1]T(t, x, y) - 2x[y^2]T(t, x, y) = tx.$$

A direct calculation gives

$$[y^1]T(t, x, y) - 2x[y^2]T(t, x, y) = x \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n} \binom{3n}{n} R^{n+1},$$

which is precisely  $xt$ , by definition (4.8) of the series  $R$ .

We have thus proved the announced expressions of the series  $\mathcal{P}$ ,  $\mathcal{C}$  and  $\mathcal{D}$ , which in turn characterise the generating functions  $P$ ,  $C$  and  $D$  of patches of various types (see (4.7)). We still have to express the generating function  $Q(t)$  of quartic Eulerian orientations in terms of  $R(t)$ . The last equation of Theorem 4.15 now reads

$$\begin{aligned} Q(t) &= \frac{1}{t^2} [y^1]\mathcal{P}(t, y) - 1 \\ &= \frac{1}{t^2} \sum_{n \geq 1} \frac{1}{n+1} \binom{2n-1}{n} \binom{3n-1}{n} R^{n+1} - 1 \\ &= \frac{1}{3t^2} \sum_{n \geq 1} \frac{1}{n+1} \binom{2n}{n} \binom{3n}{n} R^{n+1} - 1 \\ &= \frac{1}{3t^2} (t - R - 3t^2) \end{aligned}$$

by definition of  $R$ . □

It remains to prove the following lemma, used in the above proof.

**Lemma 4.25.** *For any indeterminates  $u$  and  $z$ , the Catalan series  $\text{Cat}(u) = \sum_{n \geq 0} \binom{2n}{n} u^n / (n+1)$  and the series  $A(u, z)$  defined by (4.15) are related by:*

$$\exp(A(u, z)) (1 - zu \text{Cat}(u)) = 1.$$



*Proof.* Equivalently, what we want to prove reads

$$A(u, z) = \log \frac{1}{1 - zu \operatorname{Cat}(u)} = \sum_{j \geq 0} \frac{z^{j+1}}{j+1} (u \operatorname{Cat}(u))^{j+1}.$$

Comparing with the expansion in  $z$  of  $A(u, z)$  (see (4.15)), what we want to show is

$$[u^{n+1}](u \operatorname{Cat}(u))^{j+1} = \frac{j+1}{n+1} \binom{2n-j}{n}.$$

This follows from the Lagrange inversion formula [106, p. 732], applied to

$$F(u) := u \operatorname{Cat}(u) = \frac{u}{1 - F(u)}.$$

□

**Remark.** The above lemma is a special case of a general identity which relates the enumeration of two classes of one-dimensional lattice paths, sharing the same step set, both constrained to end at a non-negative position. For the first class there is no other condition, while for the second class the path is not allowed to visit any negative point. The generating functions of these two classes, counted by the number of steps (variable  $z$ ) and the final position (variable  $u$ ) are respectively denoted by  $W^+(z, u)$  and  $F(z, u)$ . Then, on p. 51 of [12], the following identity appears

$$F(z, u) = \exp \left( \int_0^z (W^+(t, u) - 1) \frac{dt}{t} \right).$$

When the only allowed steps are  $+1$  and  $-1$ , this reads, using a standard factorization on non-negative paths into *Dyck paths* (counted by  $\operatorname{Cat}(z^2)$ ):

$$\frac{\operatorname{Cat}(z^2)}{1 - zu \operatorname{Cat}(z^2)} = \exp \left( \sum_{n \geq 1} \sum_{j=0}^n \frac{1}{2n-j} \binom{2n-j}{n} z^{2n-j} u^j \right).$$

Upon dividing this identity by its specialization at  $u = 0$ , we obtain

$$\frac{1}{1 - zu \operatorname{Cat}(z^2)} = \exp \left( \sum_{n \geq 1} \sum_{j=1}^n \frac{1}{2n-j} \binom{2n-j}{n} z^{2n-j} u^j \right).$$

Now some elementary transformations (involving replacing  $u$  by  $uz$ , then  $z$  by  $\sqrt{z}$ , and finally swapping  $u$  and  $z$ ) shows that this is equivalent to our lemma.

## A bijection

In 1998, Schaeffer found a bijection between marked planar quadrangulations and planar trees [180], using the intermediate step that these are each in bijection with labelled quadrangulations with a unique local minimum. This was generalised in 2004 by Bouttier, Di Francesco and Guitter [52] to a bijection, now known as the BDG bijection, between marked general bipartite planar maps and a class of labelled trees which they call *mobiles*. Again, this is via a bijection with labelled maps with a unique local minimum. The Schaeffer bijection was generalised in a different way by Miermont [156] and then more explicitly by Ambjørn and Budd [7] to a bijection between labelled quadrangulations and a generalisation of labelled maps (where adjacent labels may differ by 0 or 1). We will call this the M-AB bijection. A restriction of the M-AB bijection sends labelled quadrangulations that have no inner face with only two labels to general labelled maps, which we have seen are in bijection with general Eulerian orientations. This bijection is one of the key steps in the proof of Theorem 4.2. In this section, we describe a bijection  $\Phi$  of Bouttier, Fusy and Guitter [53] between the labelled (bipartite) maps of Definition 4.3 and other bipartite maps which we call *mobile-maps*, which extends both the BDG bijection and the M-AB bijection. We then describe several interesting specializations of  $\Phi$ , including those mentioned above.

### From labelled maps to mobile-maps

**Definition 4.26.** A *mobile-map* is a rooted planar map with black and white vertices satisfying the following properties:

- (a) all edges join a black vertex to a white one; in particular, a mobile-map is bipartite,
- (b) all white vertices carry an integer label,
- (c) the root vertex is white and has label 1,
- (d) for each black vertex  $v$ , the labels  $\ell$  and  $m$  of two white vertices adjacent to  $v$  and consecutive in clockwise direction around  $v$  satisfy  $m \geq \ell - 1$ .

An example is shown in Figure 4.32.

We now describe a map  $\Phi$ , which, starting from a labelled map  $L$ , produces a mobile-map  $M$ . The construction takes place independently in every face of  $L$ , and in each face, *coincides with the construction of so-called mobiles in [52, Sec. 2.1]*. Still, we repeat it for the sake of completeness. The

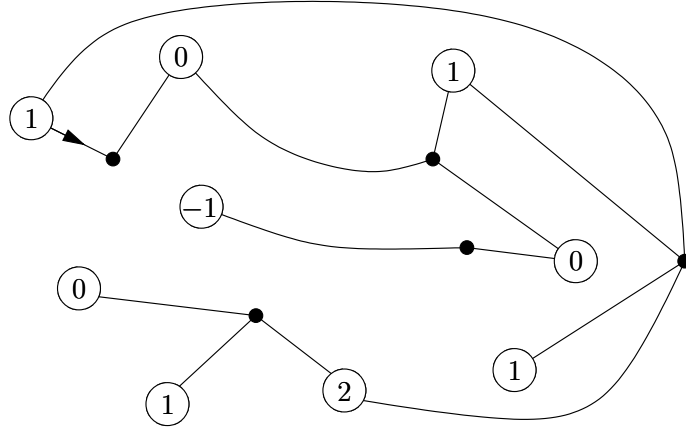


Figure 4.32: A mobile-map with 2 faces.

vertices of  $L$  are coloured white, and its edges are dashed in our figures. The construction is illustrated in Figure 4.33, left.

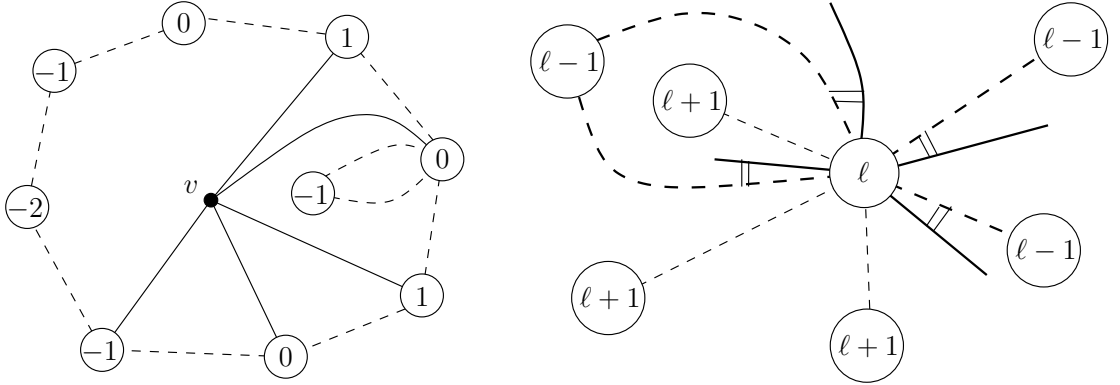


Figure 4.33: Left: construction of  $\Phi$  in a face of  $L$  of degree 10. Right: the map  $\Phi$  creates  $k$  solid edges of  $M$  at a vertex of  $L$  labelled  $\ell$  and having  $k$  incidences with neighbours labelled  $\ell - 1$  (here,  $k = 4$ ). The matching between edges of  $L$  and edges of  $M$  is emphasised.

In every face  $F$  of  $L$ , say of degree  $2k$ , we add a black vertex  $v$ . Recall that labels vary by  $\pm 1$  along edges. Hence, among the  $2k$  corners of  $F$ , exactly  $k$  are followed, in clockwise order around  $F$ , by a corner with smaller label. We join these  $k$  corners to the black vertex  $v$ , without creating edge intersections (Figure 4.33, left). We repeat this procedure in each face<sup>1</sup> of  $L$  (see Figure 4.34 for a complete example). Observe that each vertex of  $L$

<sup>1</sup>Note that going clockwise around  $F$  means going with the edges of  $F$  on the left: hence in the outer face, the construction seems to proceed anticlockwise.

that is not a local minimum is now joined to a black vertex. More generally, the number of new (solid) edges starting at a vertex of  $L$  labelled  $\ell$  is the number of edges that join it to a vertex labelled  $\ell - 1$  (Figure 4.33, right). Since the root edge is oriented from 0 to 1, the co-root vertex must be joined to the black vertex  $v$  located in the co-root face of  $L$ . We orient the edge that joins it to  $v$  towards  $v$ : this will be the root edge of the new object. Finally, we delete all edges of  $L$ , and also all vertices of  $L$  that have become isolated: they are those whose label is a local minimum. We denote by  $\Phi(L)$  the resulting object, which is a planar graph embedded in the plane, with a root edge joining a white vertex labelled 1 to a black vertex.

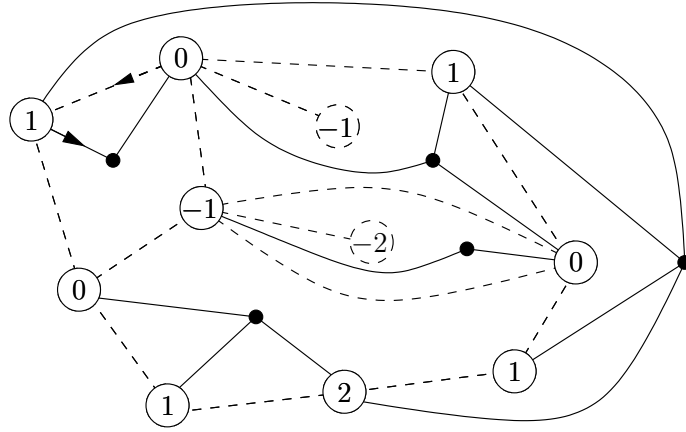


Figure 4.34: The labelled map  $L$  of Figure 4.4 (dashed edges) and the associated mobile-map  $M = \Phi(L)$  (solid edges) superimposed. The mobile-map  $M$  is that of Figure 4.32. Two white vertices of  $L$ , namely its local minima, shown in dashed disks, disappear when constructing  $M$ .

The following Proposition is essentially Theorem 1 from [53]. In their terminology, a mobile map is a star representations of a well-labelled hyper-map.

**Proposition 4.27.** *The transformation  $\Phi$  bijectively sends labelled maps to mobile-maps. Moreover, if  $\Phi(L) = M$ , then the number of vertices, edges and faces of  $L$  and  $M$  are related as follows:*

$$v_{\bullet}(M) = f(L), \quad v_{\circ}(M) = v(L) - v_{\min}(L), \quad e(M) = e(L), \quad f(M) = v_{\min}(L), \quad (4.17)$$

where  $v_{\bullet}(M)$  (resp.  $v_{\circ}(M)$ ) denotes the number of black (resp. white) vertices of  $M$ , and  $v_{\min}(L)$  denotes the number of local minima in  $L$ . Moreover, each face of degree  $2k$  in  $L$  is transformed into a black vertex of degree  $k$  in  $M$ .

*Proof.* Theorem 1 from [53] states, amongst other things, that this map  $\Phi$  is a bijection which satisfies  $v_o(M) = v(L) - v_{\min}(L)$  and  $f(M) = v_{\min}(L)$ , so we will just prove the other two equations. As we have seen, in the superposition of  $M$  and  $L$ , there are exactly  $k$  edges of  $M$  inside each face  $F$  of  $L$  with degree  $2k$ . Summing over all faces  $F$  of  $L$ , we see that the number of edges in  $M$  is equal to half of the sum of the degrees of faces of  $L$ , which is equal to the number of edges in  $L$ . Hence  $e(M) = e(L)$ . Now by Euler's formula,

$$v_{\bullet}(M) + v_o(M) + f(M) - e(M) = 2 = v(L) + f(L) - e(L).$$

Subtracting  $v_o(M) + f(M) - e(M) = v(L) - v_{\min}(L) + v_{\min}(L) - e(L)$  from both sides of this equation yields the final equation  $v_{\bullet}(M) = f(L)$ .  $\square$

## Specialization to labelled quadrangulations

We now specialise  $\Phi$  to a labelled quadrangulation  $L$ : by Proposition 4.27, every black vertex of  $\Phi(L)$  now has degree 2, and thus can be simply erased to yield a map with white vertices only. Condition (4) of the definition of mobile-maps (Definition 4.26) translates into increments  $0, \pm 1$  along edges. This yields the following proposition, which is essentially Theorem 1 from [7].

**Proposition 4.28.** *Upon erasing black vertices, the map  $\Phi$  induces a bijection  $\Phi_4$  between labelled quadrangulations and planar maps with integer labels on vertices, differing by  $0, \pm 1$  along edges, having root vertex labelled 1. Moreover, if  $\Phi_4(Q) = M$ , then*

$$e(M) = f(Q), \quad f(M) = v_{\min}(Q).$$

*The first identity can be refined as follows: a face of  $Q$  in which only two different labels occur gives rise to an edge of  $M$  with increment 0, while a face where three different labels occur gives rise to an edge with increment  $\pm 1$ .*

**Remark.** In [7], Ambjørn and Budd do not explicitly state the refinement in our proposition, although this is clear from the definition of  $\Phi$ . On the other hand, Ambjørn and Budd also prove as part of their Theorem 1 that  $Q$  and  $M$  have the same maximal label and the same number of local maxima and that the minimal label in  $M$  is one greater than the minimal label in  $Q$ .

Of particular importance will be labelled quadrangulations in which every face (including the outer one) contains three distinct labels: we call them *colourful*. Take a colourful labelled quadrangulation  $Q$ . By the above proposition, the map  $M := \Phi_4(Q)$  has all increments equal to  $\pm 1$ . Its root vertex

is labelled 1, hence the root edge is labelled either from 1 to 0, or from 1 to 2. In the former case, reversing the direction of the root edge gives a labelled map (in the sense of Definition 4.3). In the latter case, subtracting 1 from every label yields a labelled map. Conversely, take a labelled map  $L$ , and reverse the orientation of its root edge: this yields a map of the form  $\Phi_4(Q)$ , in which the root edge has labels 1 and 0 (Figure 4.35, left). Alternatively, one can add 1 to every label: the resulting map is of the form  $\Phi_4(Q')$ , and the root edge has labels 1 and 2 (Figure 4.35, right). This gives a 2-to-1 correspondence between colourful labelled quadrangulations and labelled maps.

**Corollary 4.29.** *The number of labelled colourful quadrangulations with  $n$  faces and  $k$  local minima equals twice the number of labelled maps with  $n$  edges and  $k$  faces, or equivalently, twice the number of Eulerian orientations with  $n$  edges and  $k$  vertices.*

This corollary will be the key in our enumeration of general Eulerian orientations. If we start from a general labelled quadrangulation  $Q$ , possibly containing faces with only two labels, then we can still apply the duality rule of Figure 4.5 to the map  $\Phi_4(Q)$  (carrying labels), with the additional rule that we do not orient an edge that lies between two faces with the same label. In this way we obtain a *partial* Eulerian orientation, that is, a map in which *some* edges are oriented, in such a way that there are as many incoming as outgoing edges at any vertex (Figure 4.36).

**Corollary 4.30.** *There is a bijection between quartic Eulerian orientations with  $n$  vertices (in which the root edge is oriented canonically) and partial Eulerian orientations with  $n$  edges (with no orientation requirement on the root edge).*

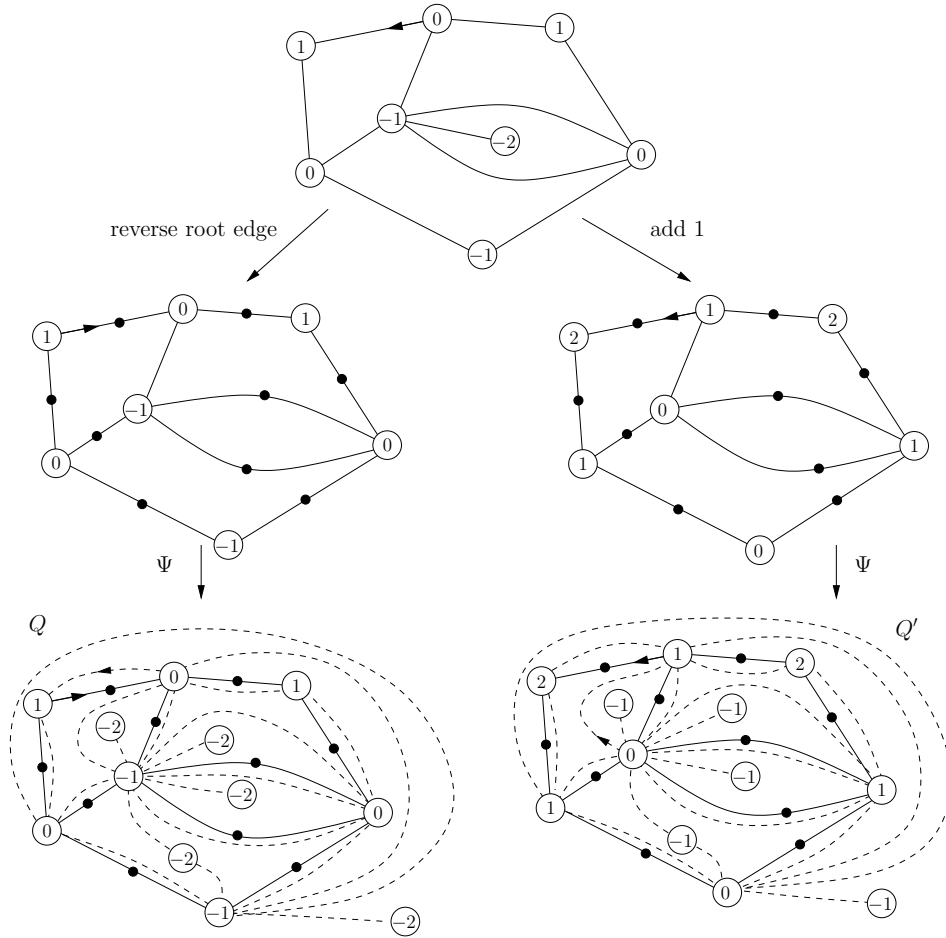


Figure 4.35: One labelled map gives two mobiles with black vertices of degree 2, which give in turn two colourful labelled quadrangulations  $Q$  and  $Q'$  (in dashed edges). They only differ by a shift of labels and a change in the root edge.

## Specialization to canonically labelled maps

Given a bipartite map, there are two obvious ways to transform it into a labelled map  $L$ :

- We could equip it with the distance labelling, in which the label of each vertex is its graph distance to the root vertex. In this case the root vertex is the only local minimum, and the mobile-map  $\Phi(L)$  is just the mobile-tree of the BDG-bijection [52, Sec. 2].
- We could bicolour all vertices with zeros and ones, assigning label 0 to the root vertex. Then the map  $\Phi$  just creates a black vertex in every face of  $L$ , and joins it to all vertices labelled 1, giving a new bipartite

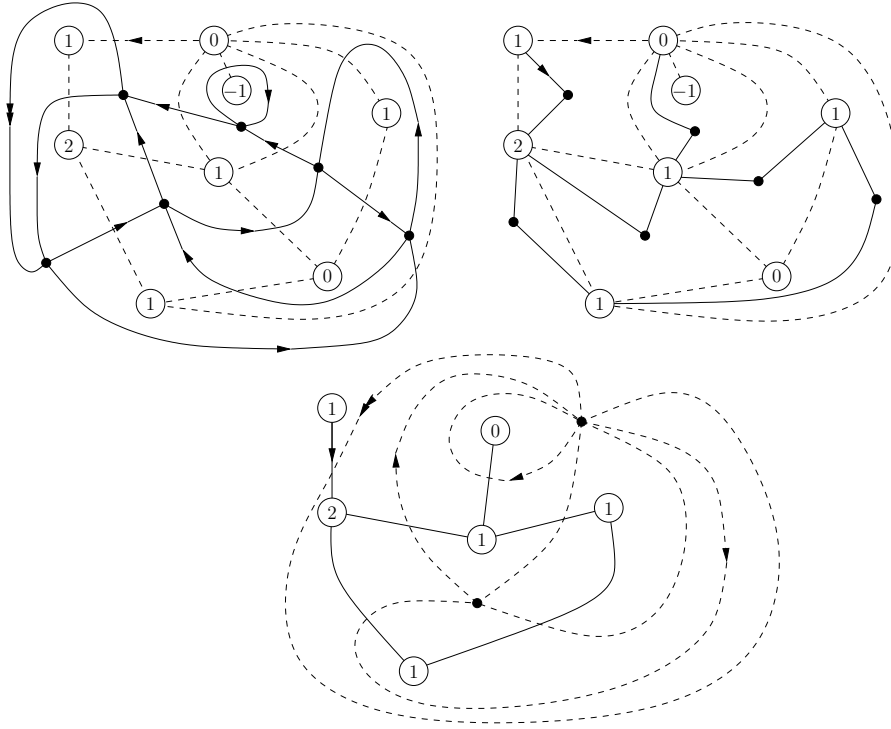


Figure 4.36: From quartic Eulerian orientations to Eulerian *partial* orientations. Top left: a quartic Eulerian orientation, shown by solid edges, and the dual labelled quadrangulation  $Q$  (dashed edges). Top right: applying  $\Phi$  to  $Q$  gives a mobile-map  $M$  with black vertices of degree 2 (solid edges). Bottom: upon erasing the black vertices of  $M$ , and re-applying duality, one obtains an Eulerian partial orientation of a new map (dashed edges).

map. If we label the black vertices of  $\Phi(L)$  with zeros, then applying  $\Phi$  to  $\Phi(L)$  gives back the original bipartite map  $L$  (up to the position of the root edge; see Figure 4.37).

## Functional equations for general Eulerian orientations

In this section we will characterise the generating function  $Q^c(t)$  of colourful labelled quadrangulations (which, by Corollary 4.29, is twice the generating function of Eulerian orientations) by a system of functional equations. As one might expect, we adapt the system of Theorem 4.15 to the colourful setting. However, the third equation and the initial conditions are simpler in the colourful case.



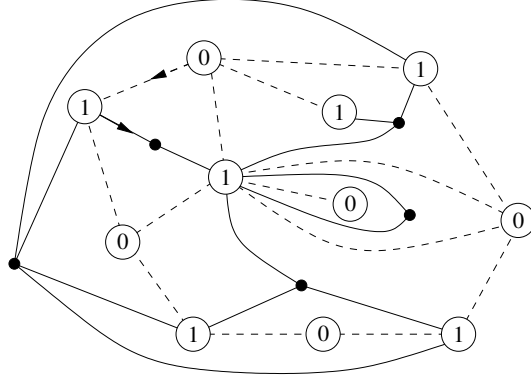


Figure 4.37: The map  $\Phi$  applied to a bipartite map canonically labelled with zeros and ones boils down to a simple involution on bipartite maps (up to the position of the root edge).

**Theorem 4.31.** *There exists a unique 3-tuple of series, denoted  $P(t, y)$ ,  $C(t, x, y)$  and  $D(t, x, y)$ , belonging respectively to  $\mathbb{Q}[[y, t]]$ ,  $\mathbb{Q}[x][[y, t]]$  and  $\mathbb{Q}[[x, y, t]]$ , and satisfying the following equations:*

$$\begin{aligned} P(t, y) &= \frac{1}{y}[x^1]C(t, x, y), \\ D(t, x, y) &= \frac{1}{1 - C\left(t, \frac{1}{1-x}, y\right)}, \\ C(t, x, y) &= xy[x^{\geq 0}] \left( P(t, tx) D\left(t, \frac{1}{x}, y\right) \right), \end{aligned}$$

together with the initial condition  $P(t, 0) = 1$ .

The generating function that counts colourful labelled quadrangulations by faces is

$$Q^c(t) = [y^1]P(t, y) - 1.$$

By Corollary 4.29,  $Q^c(t) = 2G(t)$ , where  $G(t)$  counts Eulerian orientations by edges.

**Remark.** As with the system of Theorem 4.15, the conditions on the series  $P$ ,  $C$  and  $D$  make the operations that occur in the above equations well defined. The extraction of the coefficient of  $x^1$ , and the replacement of  $x$  by  $1/(1 - x)$ , are justified as for the previous system. In the third equation, the term  $P(t, tx)D\left(t, \frac{1}{x}, y\right)$  must be seen as a power series in  $t$  and  $y$  whose coefficients are Laurent series in  $1/x$ . The extraction of the non-negative part in  $x$  then yields an element of  $\mathbb{Q}[x][[y, t]]$ .

As before, the series  $P$ ,  $C$  and  $D$  of Theorem 4.31 count certain labelled maps. Recall the definition of patches, C-patches and D-patches (Definition 4.16). Generalizing the definition of colourful quadrangulations introduced in Section 4.7.2, we say that a patch (or a D-patch) is *colourful* if each inner quadrangle contains 3 distinct labels. We define  $P(t, y)$ ,  $C(t, x, y)$  and  $D(t, x, y)$  to be respectively the generating functions of colourful patches, colourful C-patches and colourful D-patches, where  $t$  counts inner quadrangles,  $y$  the outer degree (halved), and  $x$  either the degree of the root vertex (for C-patches) or the number of inner digons (for D-patches). The equations

$$\begin{aligned} P(t, y) &= \frac{1}{y} [x^1] C(t, x, y) \\ D(t, x, y) &= \frac{1}{1 - C\left(t, \frac{1}{1-x}, y\right)}, \\ Q^c(t) &= [y^1] P(t, y) - 1, \end{aligned}$$

have identical proofs to those in Section 4.5, except that the patches, D-patches and quadrangulations in the proofs are restricted to being colourful (see Lemmas 4.17, 4.18 and 4.23).

**Remark.** The third equation of Theorem 4.15,

$$D(t, x, y) = 1 + y [x^{\geq 0}] \left( D(t, x, y) \left( \frac{1}{x} P\left(t, \frac{t}{x}\right) + [y^1] D(t, x, y) \right) \right) \quad (4.18)$$

also holds in the colourful setting, with the same proof as before (because in the proof of Lemma 4.22, all quadrangles that come from digons are automatically colourful). Its natural complement, which is the fourth equation of Theorem 4.15 (the initial condition) has no clear colourful counterpart: indeed, the relabelling of vertices that we use in Lemma 4.19, and more precisely in the fourth case of Figure 4.27, transforms the colourful quadrangles incident to the root vertex into bicoloured quadrangles (and vice versa). We could instead use the initial condition  $[y^1] C(t, x, y) = x P(t, tx)$ , which can be proved by taking a colourful C-patch of outer degree 2 and deleting the root vertex and all incident edges, then decreasing each label by 1 (Figure 4.38). However, the third equation of Theorem 4.31 is simpler than (4.18), and also relies on a simpler construction.

In order to prove the third equation of Theorem 4.31, we need some analogues of minus-patches from Section 4.5, which we call *shifted patches*.

**Definition 4.32.** A *shifted patch* is a map obtained from a patch by replacing each label  $\ell$  with  $\ell + 1$ .

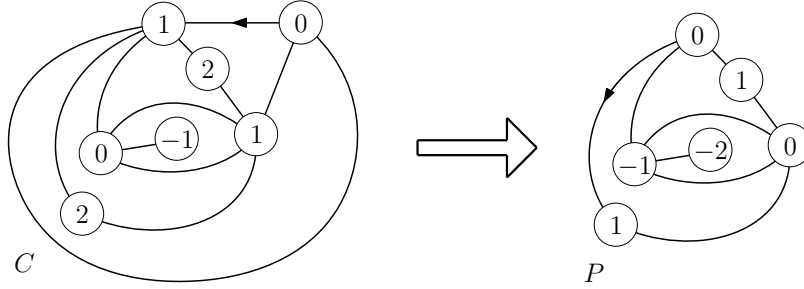


Figure 4.38: A colourful C-patch  $C$  with outer degree 2 and the corresponding patch  $P$ .

We now describe a way to extract a shifted subpatch from a patch, which parallels the extraction of a minus-patch of Definition 4.21. One minor difference is that we do not need conditions on the neighbours of the root vertex, so that we define shifted subpatches for any patch (although we will only extract them from C-patches later).

**Definition 4.33.** Let  $P$  be a patch and let  $c$  be an outer corner of  $P$  at a vertex  $v$  labelled 1. We define the *shifted subpatch of  $P$  rooted at  $c$*  as follows. First, let  $S'$  be the maximal connected submap of  $P$  that contains  $v$  and consists of vertices labelled 1 or more. Let  $S$  be the submap of  $P$  that contains  $S'$  and all edges and vertices within its boundary (assuming the root face is drawn as the infinite face). The map  $S$ , which we root at the corner inherited from  $c$ , is the *shifted subpatch of  $P$  rooted at  $c$* .

An example is shown on the left of Figure 4.39, where the shifted subpatch  $S$  is drawn with thick lines. It is easily shown that  $S$  is, as it should be, a shifted patch. The argument is the same as for minus-subpatches (in that case, the condition on the neighbours of the root having labels 1 was there to prevent the minus-subpatch to absorb the root vertex; this cannot happen with the shifted subpatch, whose boundary only contains positive labels). Every edge in  $P$  that connects a vertex in  $S$  to a vertex not in  $S$  must have endpoints labelled 1 (in  $S$ ) and 0 (not in  $S$ ), and conversely every vertex labelled 1 on the boundary of  $S$  is joined to a vertex labelled 0 out of  $S$ . We can contract  $S$  into a single vertex  $v_1$  labelled 1. This vertex is only adjacent to vertices labelled 0 in the resulting map  $L$ , and the number of digons incident to  $v_1$  is half the outer degree of  $S$ . The outer degrees of  $L$  and  $P$  coincide, because no edge of the boundary of  $P$  has been contracted.

As in the case of minus-subpatches, we can uniquely reconstruct the patch  $P$  and its marked corner  $c$  if we are given the shifted patch  $S$  and the contracted map  $L$ , together with its outer corner inherited from  $c$ . The idea is

again to attach the edges incident to  $v_1$  in  $L$  around the shifted patch  $S$ , as illustrated (in the case of minus-patches) in Figure 4.29.

We are now ready to prove the third equation of Theorem 4.31.

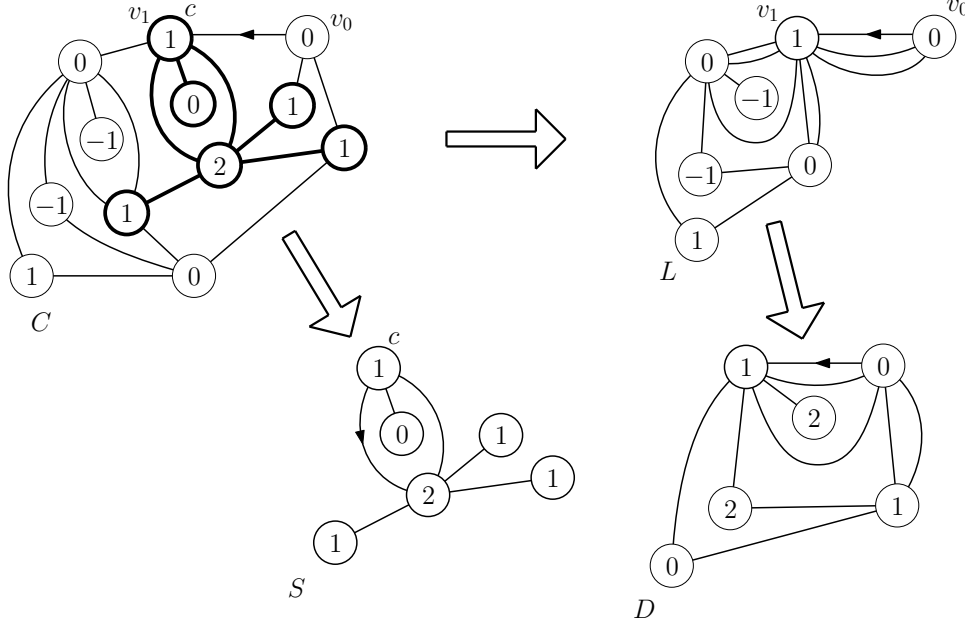


Figure 4.39: On the left is a C-patch  $C$ . The other maps shown are the shifted subpatch  $S$ , the contracted map  $L$  and the D-patch  $D$  obtained by deleting  $v_0$ , as described in the proof of Lemma 4.34.

**Lemma 4.34.** *The generating functions  $P$ ,  $C$  and  $D$  satisfy the equation*

$$C(t, x, y) = xy[x^{\geq 0}] \left( P(t, tx) D \left( t, \frac{1}{x}, y \right) \right).$$

*Proof.* Let  $C$  be any colourful C-patch. Let  $v_0$  and  $v_1$  be the root vertex and co-root vertex of  $C$ , and let  $c$  be the outer corner of  $v_1$  that is immediately anticlockwise of the root edge (we refer to Figure 4.39 for an illustration). Let  $S$  be the shifted subpatch of  $C$  rooted at  $c$  and let  $L$  be the labelled map obtained from  $C$  by contracting the subpatch  $S$  to a single vertex, still denoted  $v_1$ . Then in  $L$ , the root vertex  $v_0$  is only adjacent to vertices labelled 1 (because this was already true in  $C$ ), and the co-root vertex is only adjacent to vertices labelled 0 (because of the contraction).

Recall that all inner faces of  $L$  are either digons or quadrangles. We want to prove that in  $L$ ,  $v_0$  is not incident to any inner quadrangle. Assume that such a quadrangle exists. Since  $v_0$  only shares one corner with the outer face

of  $L$  (this was the case for  $C$  already), one such quadrangle must be incident to  $v_1$ . But the above label conditions on the neighbours of  $v_0$  and  $v_1$  force this quadrangle to have labels 0 and 1 only, in  $L$  and thus in  $C$ . This contradicts the fact that  $C$  is colourful. Hence  $v_0$  is only adjacent to inner digons (and to the outer face), and since it shares only one corner with the outer face, its only neighbour in  $L$  is  $v_1$  (see Figure 4.39).

Let  $D$  be the labelled map constructed from  $L$  by moving the root edge anticlockwise one place around the outer face, removing the old root vertex  $v_0$  of  $L$  and all incident edges, and replacing each vertex label  $\ell$  with  $1 - \ell$ . Now the root vertex has label 0, and all its neighbours have label 1. The outer face still has labels 0 and 1. Each inner quadrangle of  $D$  corresponds to an inner quadrangle of  $C$ , and is therefore colourful. Hence  $D$  is a colourful D-patch.

Let  $2j$  be the outer degree of  $S$ . Then  $j$  is also the number of inner digons in  $L$ . Let  $i \leq j$  be the number of inner digons left in  $D$  after deleting  $v_0$ . Then the number of inner digons in  $L$  that are incident to  $v_0$  (and  $v_1$ ) is  $j - i$ . Therefore, the degree of  $v_0$  in both  $L$  and  $C$  is  $j - i + 1$ . Conversely, taking a colourful D-patch  $D$  with  $i$  inner digons and a colourful shifted patch  $S$  of outer degree  $2j$  with  $j \geq i$  we construct the corresponding C-patch  $C$  as follows:

- we first construct a map  $D'$  by replacing each label  $\ell$  in  $D$  with  $1 - \ell$ ,
- next we construct a map  $L$  with a new vertex  $v_0$ , joined to the root vertex of  $D'$  by  $j - i + 1$  edges,
- finally we insert  $S$  into  $L$  to create the corresponding patch  $C$  (the choice of the corner where the subpatch extraction takes place being canonical).

As already explained, the degree of the root vertex in  $C$  is  $j - i + 1$ . The outer degree of  $C$  is the outer degree of  $D$  plus 2, and the number of inner quadrangles in  $C$  is  $j$  plus the number of inner quadrangles in  $S$  and  $D$ . Hence, with the obvious notation,

$$C(t, x, y) = \sum_{\text{od}(S) \geq \text{dig}(D)} x^{\text{od}(S) - \text{dig}(D) + 1} y^{1 + \text{od}(D)} t^{\text{qu}(S) + \text{qu}(D) + \text{od}(S)},$$

and this gives the equation of the lemma, since shifted patches are counted by  $P(t, y)$ .  $\square$

*Proof of Theorem 4.31.* Given that the initial condition  $P(t, 0) = 1$  is obvious (it accounts for the atomic patch), we have now proved all functional

equations. It remains to prove that, together with the conditions on the rings that contain  $P$ ,  $C$  and  $D$ , they determine a unique 3-tuple of series. Let us denote by  $p_{j,n}$  the coefficient of  $y^j t^n$  in  $P(t, y)$ , and similarly for  $C$  and  $D$ . These quantities should be thought of as elements of  $\mathbb{Q}$  (for  $P$ ), of  $\mathbb{Q}[x]$  (for  $C$ ) and of  $\mathbb{Q}[[x]]$  (for  $D$ ). We will prove by induction on  $N \geq 0$  that  $p_{j,n}$ ,  $c_{j,n}$  and  $d_{j,n}$  are completely determined for  $j + n \leq N$  — we say *up to order  $N$* .

First take  $N = 0$ . The third equation of the system shows that  $C$  is a multiple of  $y$ . In particular,  $c_{0,0} = 0$ . The second equation then implies that  $D - 1$  is also a multiple of  $y$ . In particular,  $d_{0,0} = 1$ . Finally, the initial condition  $P(t, 0) = 1$  gives  $p_{0,0} = 1$ . Now assume that the induction hypothesis holds for some  $N \geq 0$ , and let us prove it for  $N + 1$ .

The third equation, with its factor  $y$ , allows us to determine  $C(t, x, y)$  up to order  $N + 1$ . By construction, the coefficients  $c_{j,N+1-j}$  will be polynomials in  $x$ . Then the second equation gives  $D(t, x, y)$  up to the same order. The first equation seems to raise a problem, because of the division by  $y$ . But, combined with the third equation, it reads

$$\begin{aligned} P(t, y) &= [x^0]P(t, tx)D(t, 1/x, y) \\ &= [x^0]P(t, tx) + [x^0]P(t, tx)(D(t, 1/x, y) - 1) \\ &= P(t, 0) + [x^0]P(t, tx)(D(t, 1/x, y) - 1). \end{aligned}$$

Now  $P(t, 0) = 1$  is known, and since  $D(t, 1/x, y) - 1$  is a multiple of  $y$ , knowing  $P(t, tx)$  up to order  $N$  and  $D(t, 1/x, y)$  up to order  $N + 1$  suffices to determine  $P(t, y)$  up to order  $N + 1$ . This completes our induction.  $\square$

## Solution for general Eulerian orientations

We are now about to solve the system of Theorem 4.31, thus proving, in particular, that the generating function  $G(t)$  of Eulerian orientations is indeed given by Theorem 4.2. As in Section 4.6, the third equation of the system leads us to introduce variants of the series  $P$ ,  $C$  and  $D$ , defined again by

$$\mathcal{P}(t, y) = t P(t, ty), \quad \mathcal{C}(t, x, y) = C(t, x, ty), \quad \mathcal{D}(t, x, y) = D(t, x, ty). \quad (4.19)$$

Of course, if we determine  $\mathcal{P}$ ,  $\mathcal{C}$  and  $\mathcal{D}$ , then  $P$ ,  $C$  and  $D$  are completely determined as well.

**Theorem 4.35.** *Let  $R(t) \equiv R$  be the unique formal power series with constant term 0 satisfying*

$$t = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n}^2 R^{n+1}.$$

Then the above series  $\mathcal{P}$ ,  $\mathcal{C}$  and  $\mathcal{D}$  are:

$$\mathcal{P}(t, y) = \sum_{n \geq 0} \sum_{j=0}^n \frac{1}{n+1} \binom{2n}{n} \binom{2n-j}{n} y^j \mathbf{R}^{n+1},$$

$$\mathcal{C}(t, x, y) = 1 - \exp \left( - \sum_{n \geq 0} \sum_{j=0}^n \sum_{i=0}^n \frac{1}{n+1} \binom{2n-i}{n} \binom{2n-j}{n} x^{i+1} y^{j+1} \mathbf{R}^{n+1} \right),$$

$$\mathcal{D}(t, x, y) = \exp \left( \sum_{n \geq 0} \sum_{j=0}^n \sum_{i \geq 0} \frac{1}{n+1} \binom{2n-j}{n} \binom{2n+i+1}{n} x^i y^{j+1} \mathbf{R}^{n+1} \right).$$

The generating function of Eulerian orientations, counted by edges, is

$$\mathbf{G}(t) = \frac{1}{4t^2} (t - 2t^2 - \mathbf{R}(t)).$$

**Remark.** Observe that the series  $\mathcal{C}(t, x, y)$  is symmetric in  $x$  and  $y$ . Let us give a combinatorial explanation for this, illustrated in Figure 4.40. Consider a colourful C-patch  $C$  with root vertex  $v_0$ , and form a colourful labelled quadrangulation  $C'$  by adding a vertex  $v_2$  with label 2 to the outer face of  $C$  and joining it to each outer corner of  $C$  labelled 1. The generating function  $\mathcal{C}(t, x, y) = \mathbf{C}(t, x, ty)$  then counts the possible objects  $C'$  by the number of quadrangles (variable  $t$ ), the degree of the root vertex  $v_0$  (variable  $x$ ) and the degree of the new vertex  $v_2$  (variable  $y$ ). Moreover, the object  $C'$  can be any colourful quadrangulation in which the outer face has labels 0, 1, 2, 1 and each neighbour of either  $v_0$  or  $v_2$  is labelled 1. The transformation  $\ell \mapsto 2 - \ell$  then explains why the generating function  $\mathcal{C}(t, x, y)$  is symmetric in  $x$  and  $y$ .

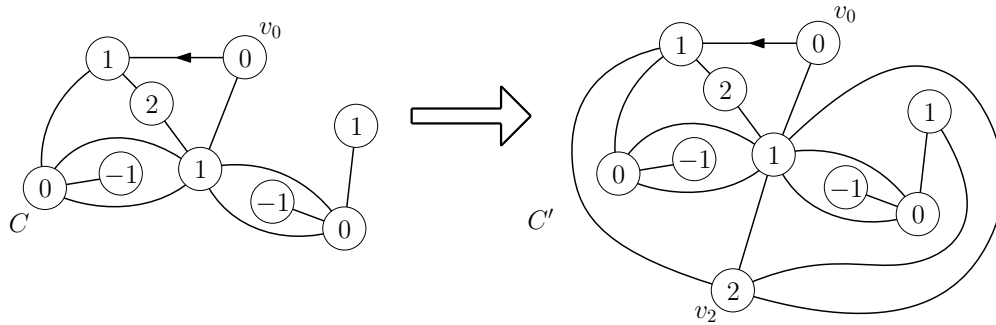


Figure 4.40: An example of the transformation from a colourful C-patch  $C$  to a colourful labelled quadrangulation  $C'$  from the remark below Theorem 4.35.

*Proof of Theorem 4.35.* We argue as in the proof of Theorem 4.24. Defining  $\mathcal{P}$ ,  $\mathcal{C}$  and  $\mathcal{D}$  as above, we first observe that the series  $\mathcal{P}$ ,  $\mathcal{C}$  and  $\mathcal{D}$  defined by (4.19) belong respectively to the rings  $\mathbb{Q}[[y, t]]$ ,  $\mathbb{Q}[x][[y, t]]$  and  $\mathbb{Q}[[x, y, t]]$ , as prescribed in Theorem 4.31. Hence it suffices to prove that they satisfy the desired system, which reads

$$\begin{aligned}\mathcal{P}(t, y) &= \frac{1}{y}[x^1]\mathcal{C}(t, x, y), \\ \mathcal{D}(t, x, y) &= \frac{1}{1 - \mathcal{C}\left(t, \frac{1}{1-x}, y\right)}, \\ \mathcal{C}(t, x, y) &= xy[x^{\geq 0}]\mathcal{P}(t, x)\mathcal{D}(t, 1/x, y), \\ \mathcal{P}(t, 0) &= t.\end{aligned}$$

Note that the first three equations do not explicitly involve the variable  $t$ : we will prove them without resorting to the definition of  $\mathbf{R}$ . But the fourth equation, namely the initial condition  $\mathcal{P}(t, 0) = t$ , does involve  $t$ , and in fact holds precisely by definition of  $\mathbf{R}$ .

The first equation is again straightforward, and the second follows from (4.12) again. Now consider the third one. Since it is more natural to handle series in  $x$  rather than in  $1/x$ , we will show instead that

$$\begin{aligned}\mathcal{C}(t, 1/x, y) &= y/x[x^{\leq 0}]\mathcal{P}(t, 1/x)\mathcal{D}(t, x, y) \\ &= [x^{< 0}](y/x \mathcal{P}(t, 1/x)\mathcal{D}(t, x, y)).\end{aligned}\tag{4.20}$$

In order to prove (4.20), we use identities that are similar to those used in the proof of (4.11). The counterpart of (4.13) is:

$$\sum_{i \geq 0} \binom{2n+i+1}{n} x^i = \frac{1}{x^{n+1}(1-x)^{n+1}} - \sum_{\ell=0}^n \binom{2n-\ell}{n} \frac{1}{x^{\ell+1}}.$$

The counterpart of (4.14) is:

$$\mathcal{D}(t, x, y) = \exp(A(\mathbf{U}, y)) (1 - \mathcal{C}(t, 1/x, y)),\tag{4.21}$$

with  $\mathbf{U} = \frac{\mathbf{R}}{x(1-x)}$  and  $A(u, y)$  is still given by (4.15). This is indeed an analogue of (4.14), since

$$1 - \mathcal{C}(t, 1/x, y) = \exp\left(-\sum_{n \geq 0} \sum_{j=0}^n \sum_{i=0}^n \frac{1}{n+1} \binom{2n-i}{n} \binom{2n-j}{n} \frac{1}{x^{i+1}} y^{j+1} \mathbf{R}^{n+1}\right)$$

can be written as  $\exp(-B(\mathbf{R}, 1/x, y))$  where  $B(\mathbf{R}, 1/x, y)$  only involves negative powers of  $x$ . By extracting the coefficient of  $y$  from (4.21), we find the



counterpart of (4.16):

$$[y]\mathcal{D}(t, x, y) = \mathbf{U} \text{Cat}(\mathbf{U}) - \frac{1}{x} \mathcal{P}\left(t, \frac{1}{x}\right), \quad (4.22)$$

where  $\text{Cat}(u)$  is still the Catalan series  $\sum_{n \geq 0} \frac{u^n}{n+1} \binom{2n}{n}$ .

With these identities at hand, we can now prove (4.20):

$$\begin{aligned} [x^{<0}] (y/x \mathcal{P}(t, 1/x) \mathcal{D}(t, x, y)) \\ &= [x^{<0}] (y \mathcal{D}(t, x, y) (\mathbf{U} \text{Cat}(\mathbf{U}) - [y]\mathcal{D}(t, x, y))) \quad \text{by (4.22),} \\ &= [x^{<0}] (y \mathcal{D}(t, x, y) \mathbf{U} \text{Cat}(\mathbf{U})) \\ &= [x^{<0}] (-\mathcal{D}(t, x, y)(1 - y \mathbf{U} \text{Cat}(\mathbf{U}))) \\ &= [x^{<0}] (-\mathcal{D}(t, x, y) \exp(-A(\mathbf{U}, y))) \quad \text{by Lemma 4.25,} \\ &= [x^{<0}] (-1 + \mathcal{C}(t, 1/x, y)) \quad \text{by (4.21),} \\ &= \mathcal{C}(t, 1/x, y). \end{aligned}$$

We have thus proved the announced expressions of  $\mathcal{P}$ ,  $\mathcal{C}$  and  $\mathcal{D}$ , which in turn characterise the generating functions  $\mathbf{P}$ ,  $\mathbf{C}$  and  $\mathbf{D}$  of colourful patches of various types. Now the last equation of Theorem 4.31 gives

$$\begin{aligned} 2\mathbf{G}(t) = \mathbf{Q}^c(t) &= \frac{1}{t^2} [y^1] \mathcal{P}(t, y) - 1 \\ &= \frac{1}{t^2} \sum_{n \geq 1} \frac{1}{n+1} \binom{2n}{n} \binom{2n-1}{n} \mathbf{R}^{n+1} - 1 \\ &= \frac{1}{2t^2} \sum_{n \geq 1} \frac{1}{n+1} \binom{2n}{n}^2 \mathbf{R}^{n+1} - 1 \\ &= \frac{1}{2t^2} (t - \mathbf{R} - 2t^2). \end{aligned}$$

The expression given in Theorem 4.35 (and in Theorem 4.2) for  $\mathbf{G}(t)$  follows.  $\square$

## Nature of the series and asymptotics

### Nature of the series

We begin by proving that the series  $\mathbf{Q}(t)$  and  $\mathbf{G}(t)$  that count respectively quartic and general Eulerian orientations satisfy non-linear differential equations of order 2, as claimed in Theorems 4.1 and 4.2. Both series are expressed in terms of a series  $\mathbf{R}$  that satisfies

$$\Omega(\mathbf{R}) = t,$$

for some hypergeometric series  $\Omega$ . In the quartic case (Theorem 4.1),

$$\Omega(r) = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n} \binom{3n}{n} r^{n+1} \quad (4.23)$$

satisfies

$$6\Omega(r) + r(27r - 1)\Omega''(r),$$

from which we derive that

$$R(27R - 1)R'' = 6tR^3.$$

Using  $3t^2Q(t) = t - 3t^2 - R(t)$ , this indeed gives a second order DE for  $Q(t)$ .

For general Eulerian orientations (Theorem 4.2), we still have  $\Omega(R) = t$ , with

$$\Omega(r) = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n}^2 r^{n+1} \quad (4.24)$$

satisfying

$$4\Omega(r) + r(16r - 1)\Omega''(r),$$

from which we derive that

$$R(16R - 1)R'' = 4tR^3.$$

Using  $4t^2G(t) = t - 2t^2 - R(t)$ , this indeed gives a second order DE for  $G(t)$ .

The fact that neither  $Q(t)$  nor  $G(t)$  solve a non-trivial *linear* DE will follow from the asymptotic behaviour of their coefficients, established in the next subsection: indeed, the logarithm occurring at the denominator prevents this behaviour from being that of the coefficients of a D-finite series [106, p. 520 and 582].

We can also describe the nature of the multivariate series counting patches.

**Proposition 4.36.** *The generating functions  $P(t, y)$ ,  $C(t, x, y)$  and  $D(t, x, y)$  counting patches of various types, and expressed in Theorem 4.24 through the identities (4.7), are D-algebraic. The same holds for their colourful counterparts, expressed in Theorem 4.35.*

*Proof.* This follows by composition of D-algebraic series (see, e.g., [25, Prop. 29]).  $\square$

Note that both series  $\mathcal{P}(t, y)$  (in the general and colourful cases) are even D-finite *as functions of  $y$  and  $R$* . The other two series  $\mathcal{C}(t, x, y)$  and  $\mathcal{D}(t, x, y)$  are clearly D-algebraic as functions of  $x, y$  and  $R$ , and it is natural to wonder if they might be D-finite. After all, in Lemma 4.25 we have met a series

that is written as the exponential of a hypergeometric series and is not only D-finite, but even algebraic.

In the one-variable setting, it is known that if  $F(t)$  is D-finite, then  $\exp(\int F(t))$  is D-finite if and only if  $F(t)$  is in fact algebraic [184]. We can use this criterion to prove, for instance, that the series  $\mathcal{D}(t, 0, 1)$  of Theorem 4.24 is not D-finite as a function of  $\mathbf{R}$ . Indeed,  $\mathcal{D}(t, 0, 1) = D(\mathbf{R})$  with

$$\begin{aligned} D(r) &= \exp \left( \sum_{n \geq 0} \sum_{j=0}^n \frac{1}{n+1} \binom{2n-j}{n} \binom{3n-j+1}{2n-j} r^{n+1} \right) \\ &= \exp \left( \sum_{n \geq 0} \frac{3n+2}{2(n+1)^2} \binom{2n}{n} \binom{3n+1}{2n} r^{n+1} \right) \\ &= \exp \left( \int F(r) \right) \end{aligned}$$

where

$$F(r) = \sum_{n \geq 0} \frac{3n+2}{2(n+1)} \binom{2n}{n} \binom{3n+1}{2n} r^n.$$

Then  $D(r)$  is D-finite if and only if  $F(r)$  is algebraic. But this is not the case, as the coefficient of  $r^n$  in  $F(r)$  is asymptotic to  $c 27^n/n$ , which reveals a logarithmic singularity in  $F(r)$  (see [103]). The same argument proves that  $\mathcal{C}(t, 1, 1)$  is not a D-finite function of  $\mathbf{R}$ .

In the colourful case (Theorem 4.35), we have

$$\mathcal{D}(t, 0, 1) = \exp \left( \sum_{n \geq 0} \frac{1}{n+1} \binom{2n+1}{n}^2 \mathbf{R}^{n+1} \right),$$

and a similar asymptotic argument proves that this cannot be a D-finite function of  $\mathbf{R}$ . The same holds for  $\mathcal{C}(t, 1, 1)$ .

## Asymptotics

As mentioned in the introduction, the series  $\mathbf{R}$  of Theorem 4.1 already occurred in the map literature, more precisely in the enumeration of quartic maps equipped with a spanning forest [48]. Its singular structure has been studied in detail, and the first part of the following result is the case  $u = -1$  of [48, Prop. 8.4]. As in [106, Def. VI.1, p. 389], we call  $\Delta$ -domain of radius  $\rho$  any domain of the form

$$\{z : |z| < r, z \neq \rho \text{ and } |\operatorname{Arg}(z - \rho)| > \phi\}$$

for some  $r > \rho$  and  $\phi \in (0, \pi/2)$ .

**Proposition 4.37.** *The series  $R$  of Theorem 4.1 has radius  $\rho = \frac{\sqrt{3}}{12\pi}$ . It is analytic in a  $\Delta$ -domain of radius  $\rho$ , and the following estimate holds in this domain, as  $t \rightarrow \rho$ :*

$$R(t) - \frac{1}{27} \sim \frac{1}{6} \frac{1 - t/\rho}{\log(1 - t/\rho)}.$$

Consequently, the  $n$ th coefficient of  $R$  satisfies, as  $n \rightarrow \infty$ ,

$$r_n := [t^n]R \sim -\frac{1}{6} \frac{\mu^n}{n^2 \log^2 n}$$

with  $\mu = 1/\rho = 4\sqrt{3}\pi$ .

Observe that this provides the asymptotic behaviour of the numbers  $q_n$  of Theorem 4.1 since  $q_n = -r_{n+2}/3$ . The correspondence between the singular behaviour of  $R(t)$  near its dominant singularity  $\rho$  and the asymptotic behaviour of its coefficients relies on Flajolet and Odlyzko's *singularity analysis* of generating functions [105, 106]. The singular behaviour of  $R$  near  $\rho$  is obtained using the inversion relation  $\Omega(R(t)) = t$ , where the series  $\Omega$ , given by (4.23), has radius  $1/27$  and satisfies

$$\Omega\left(\frac{1}{27}(1 - \varepsilon)\right) = \frac{\sqrt{3}}{12\pi} + \frac{\sqrt{3}}{54\pi}\varepsilon \log \varepsilon + O(\varepsilon)$$

as  $\varepsilon \rightarrow 0$ .

For general Eulerian orientations, we have a similar result.

**Proposition 4.38.** *The series  $R$  of Theorem 4.2 has radius  $\rho = \frac{1}{4\pi}$ . It is analytic in a  $\Delta$ -domain of radius  $\rho$ , and the following estimate holds in this domain, as  $t \rightarrow \rho$ :*

$$R(t) - \frac{1}{16} \sim \frac{1}{4} \frac{1 - t/\rho}{\log(1 - t/\rho)}.$$

Consequently, the  $n$ th coefficient of  $R$  satisfies, as  $n \rightarrow \infty$ ,

$$r_n := [t^n]R \sim -\frac{1}{4} \frac{\mu^n}{n^2 \log^2 n}$$

with  $\mu = 1/\rho = 4\pi$ .

As above, this gives the asymptotic behaviour of the numbers  $g_n$  of Theorem 4.1 since  $g_n = -r_{n+2}/4$ .

The proof closely follows the proof of Proposition 4.37 given in [48, Sec. 8.3], and we will not give any details. The series  $\Omega$  is now given by (4.24), has radius of convergence  $1/16$ , and satisfies

$$\Omega\left(\frac{1}{16}(1 - \varepsilon)\right) = \frac{1}{4\pi} + \frac{1}{16\pi}\varepsilon \log \varepsilon + O(\varepsilon).$$

One key ingredient is that  $t - R(t)$  has non-negative coefficients, which simply follows from the fact that this series equals  $2t^2 + 4t^2G(t)$ , by Theorem 4.2.

## Accuracy of the empirical analysis

Now that we know the exact forms of the generating functions  $Q(t)$  and  $G(t)$ , we can comment on the accuracy of the empirical analysis in section 4.3.5. Indeed, the predictions of the growth rates, and the asymptotic forms of the series were correct, and these predictions were a vital steps towards our discovery of the exact solutions. Recall that the first 100 coefficients of  $Q(t)$  were used to predict the subsequent 1000 ratios using the method of differential approximants. Using the exact solution, we have now calculated the first 998 coefficients of  $Q(t)$ . The final ratio that we have calculated is

$$\frac{[t^{998}]Q(t)}{[t^{997}]Q(t)} = 21.718170986407648634371728755726 \dots$$

This ratio was predicted, in Section 4.3.5, to be approximately

$$\frac{[t^{998}]Q(t)}{[t^{997}]Q(t)} \approx 21.718170986407648634371728755826.$$

So, remarkably, this ratio was predicted correctly to 29 significant digits.

## Final comments and perspectives

We have exactly solved the problem of counting planar Eulerian orientations, both in the general and in the quartic case. Our proof, based on a guess-and-check approach, should not remain the only proof. One should seek a better combinatorial understanding of our results. Can one explain why the series  $R$  of Theorem 4.1 also appears in the enumeration of quartic maps  $M$  weighted by their Tutte polynomial  $T_M(0, 1)$ ? Can one explain the forms of the series  $C$  and  $D$  in Theorems 4.24 and 4.35? What about more general vertex degrees? Can one interpolate between the results of Theorems 4.1 and 4.2, given that the second also counts a subclass of quartic Eulerian orientations (those with no alternating vertex)? In this final section we discuss the quest for bijections, and some aspects of interpolation.

## Bijections

Our results reveal an unexpected connection between Eulerian orientations of quartic maps (counted by the specialization  $|T(0, -2)|$  of their Tutte polynomial if we do not force the orientation of the root edge [211, Sec. 3.6]) and

the specialization  $T(0, 1)$  of slightly larger maps. Let us be more precise: one of the series considered in [48] is

$$F(t) = \sum_{M \text{ quartic}} t^{f(M)} T_M(0, 1),$$

which, in the classical interpretation of the Tutte polynomial [32, 202], counts quartic maps  $M$  equipped with an *internally inactive* spanning tree. Observe that  $t$  records here the number of faces, which exceeds the number of vertices by 2. Then it is proved that

$$F'(t) = 4 \sum_{i \geq 1} \frac{1}{i+1} \binom{3i}{i-1} \binom{2i+1}{i} R^{i+1},$$

where  $R$  is the series of Theorem 4.1. There is also an interpretation of  $F(t)$  in terms of spanning forests rather than spanning trees, but then some forests have a negative contribution:

$$F(t) = \sum_{\substack{M \text{ quartic} \\ F \text{ forest}}} t^{f(M)} (-1)^{c(F)-1}$$

where  $c(F)$  is the number of connected components of the forest  $F$ . One of the advantages of this description in terms of forests is that it gives a direct interpretation of  $t - R$ . Indeed, if we restrict the summation to forests not containing the root edge, then we obtain a new series, denoted  $H(t)$  in [48], which satisfies

$$H'(t) = 2(t - R).$$

Comparing with Theorem 4.1 leads to the following statement: the number of Eulerian orientations of quartic maps with  $n$  faces is  $(n+1)/6$  times the (signed) number of quartic maps with  $n+1$  faces equipped with a spanning forest not containing the root edge, every forest  $F$  being weighted by  $(-1)^{c(F)-1}$ . This is illustrated in Figure 4.41 for  $n = 3$ .

There is also an interpretation (and generalization) of  $t - R(t)$  in terms of certain trees [48, Sec. 5.1]. It involves a parameter  $u$ , which is  $-1$  for our series  $R$ . In the forest setting,  $u$  counts the number of connected components (minus 1).

**Proposition 4.39.** *Consider rooted plane ternary trees with leaves of two colours (say black and white, see Figure 4.42). Define the charge of such a tree to be the number of white leaves minus the number of black leaves. Call a tree of charge 1 balanced. Then the series  $t - R(t)$  of Theorem 4.1 counts,*

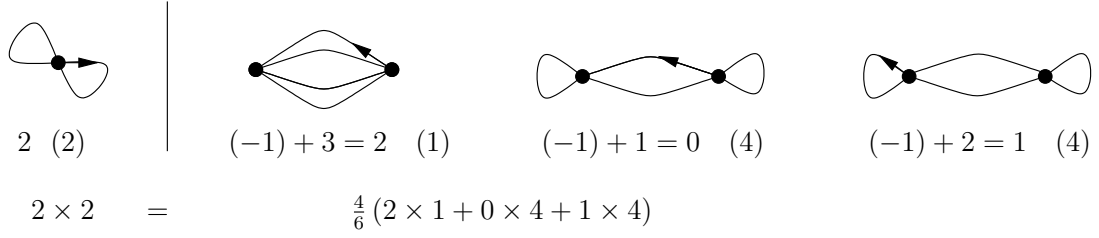


Figure 4.41: Left: the only rooted quartic graph with 1 vertex has 2 Eulerian orientations (the orientation of the root is forced) and 2 embeddings as a rooted map (with 3 faces). Right: the three rooted quartic graphs with 2 vertices, shown with the (signed) number of spanning forests avoiding the root edge. The number of embeddings as rooted planar maps is shown between parentheses.

by the number of white leaves, balanced trees in which no proper subtree is balanced.

More generally, let  $R(t, u) \equiv R$  be the only power series in  $t$  with constant term 0 satisfying

$$R = t + u \sum_{n \geq 1} \frac{1}{n+1} \binom{2n}{n} \binom{3n}{n} R^{n+1},$$

so that  $R(t, -1) = R(t)$ . Then  $(R - t)/u$  counts balanced trees by the number of white leaves, with an additional weight  $(u + 1)$  per proper balanced subtree.

Here, a *subtree* of a tree  $T$  consists of a vertex of  $T$  and all its descendants, and is *proper* if the chosen vertex is neither the root of  $T$  nor a leaf. This proposition is illustrated in Figure 4.42.

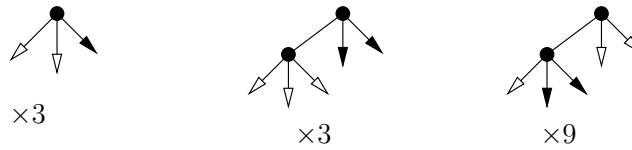


Figure 4.42: The trees with 2 and 3 white leaves involved in the expansion of  $t - R = 3t^2 + 12t^3 + O(t^4)$ , for the series  $R$  of Theorem 4.1. The multiplicities indicate the number of embeddings in the plane.

In the case of general Eulerian orientations (Theorem 4.2), we also have a similar combinatorial interpretation and generalization of  $t - R(t)$ .

**Proposition 4.40.** Consider rooted plane binary trees with edges of two types (say solid and dashed, see Figure 4.43). Define the charge of such a tree to be the number of solid edges minus the number of dashed edges. Call a tree

of charge 0 balanced. Then the series  $t - R(t)$  of Theorem 4.2 counts, by leaves, balanced trees in which no proper subtree is balanced.

More generally, let  $R(t, u) \equiv R$  be the only power series in  $t$  with constant term 0 satisfying

$$R = t + u \sum_{n \geq 1} \frac{1}{n+1} \binom{2n}{n}^2 R^{n+1}, \quad (4.25)$$

so that  $R(t, -1) = R(t)$ . Then  $(R - t)/u$  counts balanced trees by the number of leaves, with an additional weight  $(u + 1)$  per proper balanced subtree.

This proposition is illustrated in Figure 4.43.

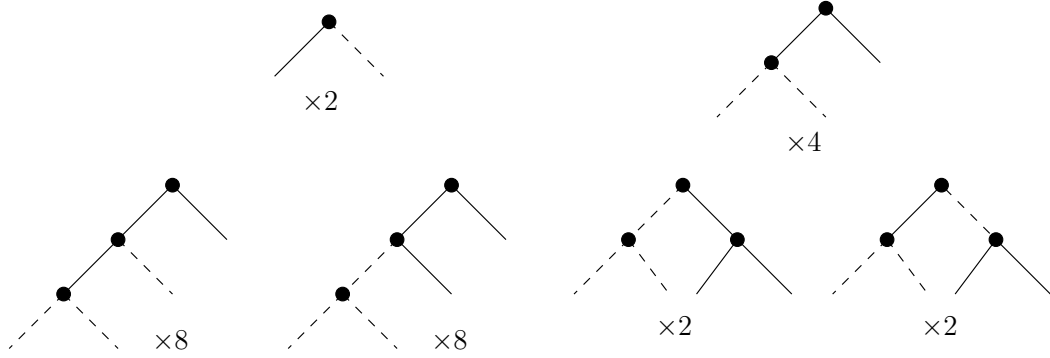


Figure 4.43: The trees with at most 4 leaves involved in the expansion of  $t - R = 2t^2 + 4t^3 + 20t^4 + O(t^5)$ , for the series  $R$  of Theorem 4.2. The multiplicities takes into account the number of embeddings in the plane and the exchange of the two edge types.

*Proof.* Define a *marked* balanced tree as a balanced tree in which a number of inner vertices are marked, in such a way that:

- the root vertex is marked (unless the tree consists of a single vertex)
- the subtree attached at any marked vertex is balanced.

Let  $\bar{R}(t, u)$  be the generating function of marked balanced trees with a weight  $t$  per leaf and a weight  $u$  per marked vertex. We claim that  $\bar{R}$  satisfies (4.25). Indeed, take a marked balanced tree with at least one edge, and consider the tree obtained by deleting all subtrees attached to a (non-root) marked vertex. Then this tree must be balanced. If it has  $n$  inner vertices, it can be chosen (removed “and coloured”) in

$$\frac{1}{n+1} \binom{2n}{n} \binom{2n}{n}$$



ways: the Catalan number accounts for the choice of the tree, and the second binomial coefficient for the ways of choosing the types of its  $2n$  edges. To reconstruct the marked balanced tree, we now need to attach to each of the  $n + 1$  leaves a marked balanced tree, and this gives (4.25). Hence the series  $\bar{R}(t, u)$  coincides with  $R(t, u)$ .

Now consider a balanced tree. The total weight of all marked trees that can be constructed from it by marking certain vertices is  $u(u + 1)^b$ , where  $b$  is the number of proper balanced subtrees. This completes the proof.  $\square$

The problem of understanding these equidistributions bijectively is wide open. Let us mention that deep connections are known to exist between certain families of orientations (e.g., acyclic) of a graph and certain families of subgraphs (e.g., spanning forests) of the same graph (see [21] for a survey, and references therein).

Let us finish with another bijective question. There exist two main bijections that transform Eulerian maps into trees: one of them takes the dual bipartite map, and transforms it into a mobile-tree using the distance labelling of the vertices [52]. This is the Bouttier–Di-Francesco–Guitter bijection that we have generalised in Section 4.7 to more general labellings, and thus to Eulerian *orientations* (rather than Eulerian maps). The second classical bijection, due to Schaeffer [179], transforms Eulerian maps into *blossoming trees*. Underlying this construction is a canonical Eulerian orientation of the map. Is there an extension of this bijection to all Eulerian orientations?

## Interpolating between quartic Eulerian orientations and general Eulerian orientations

Given that the form of our solution for general Eulerian orientations is so similar to that for quartic Eulerian orientations, one may wonder whether these are two special cases of a more general series. In a forthcoming paper Bousquet-Mélou and I describe two possible ways to simultaneously generalise  $G(t)$  and  $Q(t)$ . The first series that we consider counts general Eulerian orientations by edges and vertices. This is an obvious generalisation of  $G(t)$ , which only records the number of edges. Moreover,  $Q(t)$  can be extracted from this refined generating function by directly utilising the fact that quartic Eulerian orientations form a subclass of general Eulerian orientations (those having, in a sense, many vertices). The second generalisation concerns labelled quadrangulations, and interpolates between the series  $Q(t)$  and  $Q^c(t) = 2G(t)$  by keeping track of the number of quadrangles that only

contain two labels (such quadrangles are forbidden in colourful quadrangulations). This corresponds to the six vertex model discussed in Section [4.2.3](#).

# Chapter 5

## Enumerating cogrowth numbers in Thompson's group $F$ and related groups

### Introduction

In an attempt to find compelling evidence for the amenability or otherwise of Thompson's group  $F$ , we study the cogrowth sequence of a number of infinite, finitely generated amenable groups whose asymptotics are, in most cases, partially or fully known. Using the experience and insights gained from these examples, we turn to a study of Thompson's group  $F$ .

The cogrowth sequence of a group  $G$  with finite, inverse closed generating set  $S$  is  $(c_n)_{n \in \mathbb{N}_0}$ , where  $c_n$  is the number of words  $w$  of length  $2n$  over the alphabet  $S$ , which satisfy  $w =_G 1$  i.e.  $w$  is the identity in the group  $G$ . This is also equal to the number of walks of length  $2n$  in the Cayley graph  $\Gamma(G, S)$  which start and end at some fixed vertex  $v_0$ . There are many equivalent definitions of amenability. A standard one is that a group  $G$  is amenable if it admits a left-invariant finitely additive probability measure  $\rho$ . A consequence of the Grigorchuk-Cohen [120, 67] theorem is that  $G$  is amenable if and only if its cogrowth sequence  $(c_n)_{n \in \mathbb{N}_0}$  has exponential growth rate  $\mu = |S|^2$ . In particular, Thompson's group  $F$  is amenable if and only if its cogrowth sequence has exponential growth rate 16.

We develop new, polynomial-time algorithms to generate coefficients for a number of wreath product groups, including  $\mathbb{Z} \wr \mathbb{Z}$ ,  $(\mathbb{Z} \wr \mathbb{Z}) \wr \mathbb{Z}$  and the lamplighter group  $L = \mathbb{Z}_2 \wr \mathbb{Z}$ . We also give a polynomial time algorithm for computing terms of the cogrowth sequence of the Brin-Navas group  $B$ , and an improved algorithm to compute the terms for Thompson's group

$F$ , generating the cogrowth sequences for  $B$  and  $F$  to 128 and 32 terms respectively.

The Brin-Navas group  $B$  is one of an infinite sequence of elementary amenable two-generators subgroups of Thompson's group  $F$  described by Bleak, Brin and Moore [30]. This sequence begins with the three groups  $\mathbb{Z} \times \mathbb{Z}$ ,  $\mathbb{Z} \wr \mathbb{Z}$  and  $B$ . Since these are all amenable, and have two generators (so their inverse-closed generating sets  $S$  satisfy  $|S| = 4$ ), each of these groups has a cogrowth sequence with growth rate 16.

Recently a number of authors have considered the problem of determining which groups  $G$  have a cogrowth series

$$C_G(t) = \sum_{n \geq 0} c_n t^n$$

which is D-finite [88, 165, 20]. A result of Bell and Mishna [20] is that if  $G$  is an amenable group with super-polynomial growth, then  $C_G(t)$  is not D-finite. In particular, this implies that the cogrowth series for the groups  $L$ ,  $\mathbb{Z} \wr \mathbb{Z}$ ,  $\mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z}$  and  $B$  are not D-finite. It is not known whether any of these series is D-algebraic.

Using results of Pittet and Saloff-Coste [169, 170], we prove that the cogrowth sequence  $(c_n)_{n \in \mathbb{N}_0}$  of Thompson's group  $F$  satisfies

$$c_n < 16^n \cdot \lambda^{-n^\kappa}$$

for any real numbers  $\kappa < 1$ , and  $\lambda > 1$ . We prove this same result for the cogrowth sequence of the Brin-Navas group  $B$ . That is to say, if Thompson's group  $F$  is amenable, then its asymptotics cannot contain a stretched-exponential term<sup>1</sup>. Our numerical study reveals compelling evidence for the *presence* of such a term in the asymptotics of the coefficients of  $F$ . This suggests a potential new path to a proof of non-amenability of  $F$ : if the universality class of the cogrowth sequence can be determined rigorously, it will likely prove non-amenability.

A related result due to Moore [158] is that for any finite generating set  $S \subseteq F$ , there is a constant  $C$  such that  $F\phi_{F,S}(C^n) \geq 2 \uparrow \uparrow (n-1)$  for all sufficiently large  $n$ . Like our result, this relates to amenability, as a group  $G$  is non-amenable if and only if its Følner function  $F\phi_{G,S}$  satisfies  $F\phi_{G,S}(C^n) = \infty$  for all sufficiently large  $n$ .

---

<sup>1</sup>We define *stretched exponential* more broadly than usual. It normally refers to a term of the form  $e^{-t^\beta}$ , with  $t > 0$  and  $0 < \beta < 1$ . We allow behaviour such as  $e^{-t^\beta \cdot \log^\delta t}$ , or indeed any appropriate logarithmic term. We do not have a name for sub-exponential growth of the form  $e^{-t/\log^\delta t}$ , with  $\delta > 0$  (or appropriate logarithmic function) which is the type of term that must be present in the cogrowth series of the Brin-Navas group, and indeed in Thompson's group  $F$  if it is amenable.

In the final section of this chapter we present an empirical analysis of the cogrowth sequences of the Brin-Navas group  $B$  and Thompson's group  $F$ . This analysis is based on the series which we have generated of 128 and 32 terms, respectively. For Thompson's group  $F$ , this analysis suggests that the cogrowth sequence  $(c_n)_{n \in \mathbb{N}_0}$  behaves like

$$c_n \sim c \cdot \mu^n \cdot \kappa^{n^\sigma \log^\delta n} \cdot n^g,$$

where  $\mu \approx 15$ ,  $\kappa \approx 1/e$ ,  $\sigma \approx 1/2$ ,  $\delta \approx 1/2$ , and  $g \approx -1$ . This provides two pieces of evidence that  $F$  is non-amenable. First, our estimate of  $\mu$  is a long way from 16, which is the value required for amenability. Secondly, as we have shown, the predicted sub-exponential behaviour is inconsistent with amenability.

The make up of this chapter is as follows. In Section 5.2 we give some basic definition, including those of the groups which we consider in this chapter. In Section 5.3 we describe the algorithms developed for the cogrowth sequences of the lamplighter group  $L$ ,  $\mathbb{Z} \wr \mathbb{Z}$ ,  $\mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z}$ ,  $B$  and Thompson's group  $F$ . In Section 5.4 we discuss the possible asymptotic form of the cogrowth series for Thompson's group  $F$ , and the Brin-Navas group  $B$  and prove the absence of a stretched-exponential term if  $F$  is amenable. In Section 5.5 we present an empirical analysis of the Brin-Navas group  $B$  and Thompson's group  $F$ . Section 5.6 comprises a discussion and conclusion.

## Preliminaries

In this section we will give some basic definitions, including those of the groups which we study. In particular, we define the Heisenberg group  $H$ , wreath products with  $\mathbb{Z}$ , the Brin-Navas group  $B$  and finally Thompson's group  $F$ . For each of these groups, we will compute terms of the cogrowth sequence, which will be analysed in Section 5.5.

We start with some basic definitions regarding Cayley graphs.

**Definition 5.1.** Given a group  $G$  with finite, inverse-closed generating set  $S$ , the Cayley graph  $\Gamma(G, S)$  is defined as the graph with vertex set  $G$  and an undirected edge between two vertices  $g$  and  $h$  if and only if  $g^{-1}h \in S$ .

We denote the identity of the group  $G$  by  $e$ . Note that this is also a specific vertex in  $\Gamma(G, S)$ . For vertices  $u, v$  in a Cayley graph  $\Gamma$ , let  $d(u, v)$  denote the usual graph distance between  $u$  and  $v$ . For a vertex  $u \in V(\Gamma)$  and a non-negative integer  $n$ , the ball  $B(u, n)$  of radius  $n$  around  $u$  is defined as

$$B(u, n) = \{g \in G \mid d(u, g) \leq n\}.$$

Given a group  $G$  with finite, inverse closed generating set  $S$ , there are (at least) two natural combinatorial sequences which one can study: the growth sequence and the cogrowth sequence.

**Definition 5.2.** The *growth sequence* of  $(G, S)$  is the sequence  $(a_n)_{n \geq 0}$ , where  $a_n$  is the number of vertices  $v$  in the Cayley graph  $\Gamma(G, S)$  within distance  $n$  of  $e$ . That is,  $a_n = |B(e, n)|$ .

**Definition 5.3.** The *cogrowth sequence* of  $(G, S)$  is the sequence  $(c_n)_{n \geq 0}$ , where  $c_n$  is the number of walks of length  $2n$  in  $\Gamma(G, S)$  which start and end at  $e$ . We call walks which start and end at  $e$  *loops*.

As mentioned in the introduction, the cogrowth numbers  $c_n$  can be alternately defined without reference to the Cayley graph, as  $c_n$  is the number of words  $w$  of length  $2n$  over the alphabet  $S$  which satisfy  $\overline{w} = e$ .

Although both the growth sequence and the cogrowth sequence depend on the generating set  $S$ , some properties of these sequences are independent of the generating sets. For example, if the growth sequence of  $(G, S)$  grows exponentially, then it grows exponentially with respect to any other generating set as well. In that case the group is said to have exponential growth. Similarly a group can be said to have polynomial growth, and this is well-defined in the same sense.

It is known that a group  $G$  is amenable if and only if the exponential growth rate of the cogrowth sequence of  $(G, S)$  is equal to  $|S|^2$ . Hence this property of the cogrowth sequence is independent of the generating set.

Now we will define the groups that we study, as well as the generating sets which we will use to study them.

**Definition 5.4.** The *Heisenberg group*  $H$  is the group of matrices of the form

$$\begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$$

for integers  $a, b$  and  $c$ , under matrix multiplication.

We use the standard, inverse closed generating set  $S = \{x, y, x^{-1}, y^{-1}\}$ , where

$$x = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

The Heisenberg group has polynomial growth and its cogrowth sequence  $(c_n)_{n \in \mathbb{N}_0}$  is known [119] to behave like

$$c_n \sim \frac{1}{2} 16^n n^{-2},$$

so this serves a simple example, from both the perspective of computing terms of the cogrowth sequence and estimating the asymptotic form of this sequence.

We now define groups formed as wreath products with  $\mathbb{Z}$ . Wreath products of groups can be defined more generally, for any pair of groups, however the wreath products that we consider in this chapter all have this specific form.

**Definition 5.5.** Given a group  $G$  with finite, inverse closed generating set  $S$ , the *wreath product*  $G \wr \mathbb{Z}$  is the group whose elements are pairs  $(n, \mathbf{g})$  of an integer  $n$  and a bi-infinite sequence

$$\mathbf{g} = (g_j)_{j \in \mathbb{Z}},$$

where each  $g_j \in G$  and only finitely many elements of  $\mathbf{g}$  are not the identity  $e$  of  $G$ . For another sequence  $\mathbf{h} = (h_j)_{j \in \mathbb{Z}}$  and an integer  $m$ , the group operation is defined by

$$(n, \mathbf{g}) \cdot (m, \mathbf{h}) = (n + m, (g_j \cdot h_{j-n})_{j \in \mathbb{Z}}).$$

For an element  $g \in G$ , there is a corresponding element in  $G \wr \mathbb{Z}$ , given by  $(0, \mathbf{g})$ , where  $\mathbf{g} = (g_j)_{j \in \mathbb{Z}}$  satisfies  $g_0 = g$  and  $g_j = e$  for  $j \neq 0$ . As this forms a group embedding of  $G$  into  $G \wr \mathbb{Z}$ , we denote the element in  $G \wr \mathbb{Z}$  corresponding to  $g$  by  $g$ . We denote by  $a$  the element  $(1, (e)_{j \in \mathbb{Z}})$ . The generating set which we use for  $G \wr \mathbb{Z}$  is  $S \cup \{a, a^{-1}\}$ .

This allows us to define the lamplighter group  $L = \mathbb{Z}_2 \wr \mathbb{Z}$  as well as iterated wreath products  $\mathbb{Z} \wr_d \mathbb{Z}$  defined by  $\mathbb{Z} \wr_0 \mathbb{Z} = \mathbb{Z}$  and  $\mathbb{Z} \wr_{d+1} \mathbb{Z} = (\mathbb{Z} \wr_d \mathbb{Z}) \wr \mathbb{Z}$  for any integer  $d \geq 0$ .

**Definition 5.6.** The group  $B$  is defined as the semi-direct product

$$(\dots \wr \mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z} \wr \dots) \rtimes \mathbb{Z},$$

where the copies of  $\mathbb{Z}$  in the wreath product are generated by the elements  $\dots, a_2, a_1, a_0, a_{-1}, a_{-2}, \dots$  respectively and the generator  $s$  of the other copy of  $\mathbb{Z}$  satisfies  $sa_i s^{-1} = a_{i+1}$  for each  $i$ . To understand this infinite wreath product, note that if  $G$  is the subgroup generated by  $a_k, a_{k+1}, a_{k+2}, \dots$ , then the subgroup generated by  $G$  and  $a_{k-1}$  is  $G \wr \mathbb{Z}$ .

Note that the group  $B$  is generated by the two elements  $s$  and  $a = a_0$  because  $a_i = s^i a s^{-i}$  for each integer  $i$ , so we use the inverse closed generating set

$$S = \{a, a^{-1}, s, s^{-1}\}.$$

The group  $B$  was described independently as Exemple 6.3 in [161] and on Page 638 in [36], where Brin showed that  $B$  is an amenable subgroup of Thompson's group  $F$ . In that paper it is the group generated by  $f$  and  $h$ .

Finally, we will define Thompson's group  $F$ . The original definition given by Richard Thompson describes  $F$  as a sub-group of the group of increasing piecewise linear functions  $f : [0, 1] \rightarrow [0, 1]$  under function composition. Here we will define  $F$  using the forest representation described by Belk and Brown [19], as this will be more convenient when we describe our algorithm. Before we define Thompson's group  $F$ , we define forest diagrams.

**Definition 5.7.** A *forest diagram* is a pair  $(\mathcal{F}_1, \mathcal{F}_2)$  of sequences of planar binary trees such that each sequence has the same total number of leaves, with one tree highlighted in each sequence.

Each element of  $F$  is given by an equivalence class of forest diagrams, so to describe the elements of  $F$ , we need to describe which forest diagrams are equivalent. We will describe two types of *equivalence operation* on these forest diagrams. The first equivalence operation is the following: given a forest diagram  $(\mathcal{F}_1, \mathcal{F}_2)$  and the planar binary tree  $T$  consisting of a single vertex, the tree  $T$  can be added to the end of both sequences  $\mathcal{F}_i$  or to the start of both sequences to create an equivalent forest diagram  $(\mathcal{F}'_1, \mathcal{F}'_2)$ . Of course, this also implies that if both sequences end (or begin) with single vertex trees, then this tree can be removed from each sequence  $\mathcal{F}_i$  to create an equivalent forest diagram.

To describe the second equivalence operation, we need to define the matching of the vertices between the sequences  $\mathcal{F}_1$  and  $\mathcal{F}_2$ . Given a sequence  $\mathcal{F}_i$  of planar trees there is a unique ordering on the leaves such that the leaves in each tree are ordered in the natural way and all leaves in a single tree of the sequence  $\mathcal{F}_i$  occur to the left of all leaves in the following tree in  $\mathcal{F}_i$ . For each  $n$ , the  $n$ th leaf in  $\mathcal{F}_1$  is *matched* with the  $n$ th leaf in  $\mathcal{F}_2$ . The second equivalence operation is the following: given a pair  $(v_1, v_2)$  of matching leaves in a forest diagram, one can change each of the leaves  $v_i$  to a vertex with two children  $v_{i,0}$  (left) and  $v_{i,1}$  (right) to create an equivalent forest diagram  $(\mathcal{F}'_1, \mathcal{F}'_2)$ . Note that after this operation  $v_{1,0}$  and  $v_{2,0}$  will be matching leaves, as will  $v_{1,1}$  and  $v_{2,1}$ . Of course, this operation can be reversed: if there are vertices  $v_1$  and  $v_2$  which have children  $v_{1,0}$ ,  $v_{1,1}$ ,  $v_{2,0}$  and  $v_{2,1}$  such that  $(v_{1,0}, v_{2,0})$  and  $(v_{1,1}, v_{2,1})$  are pairs of matching leaves, then these four leaves can be removed to create an equivalent forest diagram.

We say that two forest diagrams  $(\mathcal{F}_1, \mathcal{F}_2)$  and  $(\mathcal{F}'_1, \mathcal{F}'_2)$  are *equivalent*, and write  $(\mathcal{F}_1, \mathcal{F}_2) \sim (\mathcal{F}'_1, \mathcal{F}'_2)$ , if it is possible to change one into the other using the equivalence operations. Clearly  $\sim$  is an equivalence relation, justifying



our use of the word *equivalent*. The elements of Thompson's group are equivalence classes of  $\sim$ . We call a forest diagram  $(\mathcal{F}_1, \mathcal{F}_2)$  *reduced* if neither of the two equivalence operation can be used in reverse on  $(\mathcal{F}_1, \mathcal{F}_2)$ . An important result from [19] is that each equivalence class  $g$  of  $\sim$  (that is, each  $g \in F$ ) contains exactly one reduced forest diagram.

To define the group operation, on  $F$ , we use the following fact from [19]: if  $(\mathcal{F}_1, \mathcal{F}_2)$  and  $(\mathcal{F}_3, \mathcal{F}_4)$  are forest diagrams, then there exist equivalent forest diagrams  $(\mathcal{F}'_1, \mathcal{F}'_2)$  and  $(\mathcal{F}'_3, \mathcal{F}'_4)$  satisfying  $\mathcal{F}'_2 = \mathcal{F}'_3$ . Then the operation is defined by

$$(\mathcal{F}_1, \mathcal{F}_2) \cdot (\mathcal{F}_3, \mathcal{F}_4) = (\mathcal{F}'_1, \mathcal{F}'_4).$$

We will now describe the generators  $a, b \in F$ . Let  $\mathcal{F}_\alpha$  be the sequence of two single vertex trees with the first copy highlighted, let  $\mathcal{F}_\beta$  be the sequence of two single vertex trees with the second copy highlighted and let  $\mathcal{F}_\gamma$  be the sequence containing only one tree: the binary tree with two leaves, with this tree highlighted. Then  $a$  is given by the pair  $(\mathcal{F}_\alpha, \mathcal{F}_\beta)$  while  $b$  is given by the pair  $(\mathcal{F}_\alpha, \mathcal{F}_\gamma)$ .

To check that the group operation is well defined and that  $a$  and  $b$  generate the group  $F$  is rather tedious but not difficult, so we do not include these details here. For more detail, see [19].

## Series generation

In this section we describe the algorithms we have used to compute the terms of the cogrowth sequence of various groups. We start by describing polynomial time algorithms which we have found and used for the groups  $L$ ,  $\mathbb{Z} \wr \mathbb{Z}$ ,  $\mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z}$  and  $B$ . Finally we describe the algorithm which we have used for Thompson's group  $F$ . The first 50 coefficients for the group  $B$  are given in Table 5.1, while the coefficients of the cogrowth series of  $F$  are given in Table 5.2.

## Wreath Products $G \wr \mathbb{Z}$

Let  $G$  be a group with finite, inverse closed generating set  $S$ . In this section we study the group  $G \wr \mathbb{Z}$  with respect to the generating set  $\{a, a^{-1}\} \cup S$ , where  $a$  generates the top group  $\mathbb{Z}$ . We will describe a polynomial time algorithm for computing the cogrowth series of  $G \wr \mathbb{Z}$  given the corresponding series for  $G$ . In particular, this yields a polynomial time algorithm to compute the cogrowth sequence of the lamplighter group  $\mathbb{Z}_2 \wr \mathbb{Z}$  as well as groups such as  $\mathbb{Z} \wr \mathbb{Z}$  and  $(\mathbb{Z} \wr \mathbb{Z}) \wr \mathbb{Z}$ .

Let  $a_k$  be the number of loops of length  $k$  in  $G$ . For example, if  $G = \mathbb{Z}$ , then  $a_{2k} = \binom{2k}{k}$  and  $a_{2k+1} = 0$  for all  $k \in \mathbb{Z}_{\geq 0}$ . Let  $F(t)$  be the generating function for loops in  $G \wr \mathbb{Z}$  counted by length, so the  $n$ th term of the cogrowth sequence of  $G \wr \mathbb{Z}$  is  $[t^{2n}]F(t)$ . Given a positive integer  $N$  and values  $a_k$  for  $k \leq N$  our algorithm will compute the first  $N$  coefficients of  $F(t)$ .

For each positive integer  $n$ , let  $P_n(t)$  denote the generating function for  $n$ -tuples of words  $w_1, w_2, \dots, w_n \in S^*$  such that  $\overline{w_1 \dots w_n} = 1$ , counted by the length of the word  $w_1 \dots w_n$ .

**Lemma 5.8.** *The generating function  $P_n(t)$  is given by the equation*

$$P_n(t) = \sum_{j=0}^{\infty} \binom{j+n-1}{n-1} a_j t^j.$$

*Proof.* We will calculate the coefficient  $[t^j]P_n(t)$  for some positive integer  $j$ . Let  $w_1, w_2, \dots, w_n$  be an  $n$ -tuple of words which contributes to  $[t^j]P_n(t)$ , and let  $w$  denote the word  $w_1 \dots w_n$ . So  $w$  has length  $j$ . There are  $a_j$  possible words  $w$  and each such word can be separated into  $n$  subwords in  $\binom{j+n-1}{n-1}$  ways. Hence the number of possible  $n$ -tuples  $w_1, w_2, \dots, w_n$  is given by

$$[t^j]P_n(t) = \binom{j+n-1}{n-1} a_j.$$

□

Now we will define the base loop of a loop  $l$  in  $G \wr \mathbb{Z}$ .

**Definition 5.9.** Given a loop  $l$  in  $G \wr \mathbb{Z}$ , we define the *base loop*  $l'$  of  $l$  to be the loop in  $\mathbb{Z}$  made up of only the terms  $a$  and  $a^{-1}$  in  $l$ .

For each positive integer  $i$ , let  $c_i$  be the number of steps in the base loop  $l'$  from  $a^{i-1}$  to  $a^i$  (which is the same as the number of steps from  $a^i$  to  $a^{i-1}$ ) and let  $d_i$  be the number of steps from  $a^{-i+1}$  to  $a^{-i}$ . Let  $m$  and  $n$  be maximal such that  $c_m, d_n > 0$ . Then the length of  $l'$  is given by

$$|l'| = \sum_{i=1}^m 2c_i + \sum_{j=1}^n 2d_j.$$

**Lemma 5.10.** *If  $m, n \geq 1$  then the generating function  $L_{l'}(t)$  for loops  $l$  corresponding to the base loop  $l'$ , counted by length, is given by*

$$L_{l'}(t) = t^{|l'|} P_{d_n}(t) P_{c_m}(t) P_{c_1+d_1+1}(t) \prod_{i=1}^{m-1} P_{c_i+c_{i+1}}(t) \prod_{j=1}^{n-1} P_{d_j+d_{j+1}}(t). \quad (5.1)$$

If  $m = 0$  and  $n \geq 1$  then

$$L_{l'}(t) = t^{|l'|} P_{d_n}(t) P_{d_1+1}(t) \prod_{j=1}^{n-1} P_{d_j+d_{j+1}}(t).$$

If  $m \geq 1$  and  $n = 0$  then

$$L_{l'}(t) = t^{|l'|} P_{c_m}(t) P_{c_1+1}(t) \prod_{i=1}^{m-1} P_{c_i+c_{i+1}}(t).$$

If  $m = n = 0$  then  $L_{l'}(t) = P_1(t)$ .

*Proof.* Let the baseloop  $l' = a_1 a_2 \dots a_{|l'|}$  where each  $a_i \in \{a, a^{-1}\}$  and let  $l = w_1 a_1 w_2 \dots a_{|l'|} w_{|l'|+1}$ , where each  $w_i$  is a word in  $S^*$ . We say that the height of one of the subwords  $w_i$  is equal to the integer  $p$  which satisfies  $a^p = a_1 \dots a_i$ . Then  $l$  is a loop if and only if for any height  $h$ , concatenating all of the words  $w_i$  at height  $h$  creates a loop in  $G$ . Hence the generating function for the sections at height  $h$  is  $P_r(t)$  where  $r$  is the number of these sections. If  $h > 0$  then  $r = c_h + c_{h+1}$ , if  $h < 0$  then  $r = d_{-h} + d_{-h+1}$  and if  $h = 0$  then  $r = c_1 + d_1 + 1$ . Multiplying the value  $P_r(t)$  for each value of  $r$ , then multiplying by  $t^{|l'|}$  to account for the steps in the base loop yields the desired result.  $\square$

By definition,  $F(t)$  is equal to the sum of  $L_{l'}(t)$  over all possible base loops  $l'$ .

**Lemma 5.11.** *For a given pair of sequences  $c_1, \dots, c_m, d_1, \dots, d_n$ , the number of such base loops is equal to*

$$\binom{c_1 + d_1}{c_1} \prod_{i=1}^{m-1} \binom{c_i + c_{i+1} - 1}{c_i - 1} \prod_{j=1}^{n-1} \binom{d_j + d_{j+1} - 1}{d_j - 1}. \quad (5.2)$$

*Proof.* For each  $i > 0$  let a step from  $a^i$  to  $a^{i-1}$  be called a left step, and let a step from  $a^{i-1}$  to  $a^i$  be called a right step. Then the last step from  $a^i$  must be a left, and there must be  $c_i - 1$  other left steps and  $c_{i+1}$  right steps from  $a^i$ . Hence there are  $\binom{c_i + c_{i+1} - 1}{c_i - 1}$  possible orders of the steps leaving  $a^i$ . Similarly for  $j > 0$ , there are  $\binom{d_j + d_{j+1} - 1}{d_j - 1}$  possible orders of the steps leaving  $a^{-j}$ . Finally, there are  $\binom{c_1 + d_1}{c_1}$  possible orders of the steps leaving the vertex 0. The number of possible choices of these orders is the desired result, so we just need to check that every possible choice of orders corresponds to a unique base loop  $l'$ . Given an ordering of the outgoing steps from each  $a^i$ , the

loop  $l'$  can be uniquely constructed one step at a time. Hence it suffices to show each choice of orderings does indeed form a base loop  $l'$ . Given a choice of orderings, let  $p$  be the path starting at  $a^0$  formed by adding steps one at a time according to these orderings, stopping only when there are no more outgoing steps available from the current point  $a^i$ . Then  $p$  ends at  $a^0$ , as for each other point  $a^i$ , the number of available incoming steps is equal to the number of available outgoing steps, so it is impossible to run out of outgoing steps. At  $a^0$  there are also equally many available incoming and outgoing steps, so all of these steps must have been used. Moreover, if for some  $i > 0$ , not all available outgoing steps from  $a^i$  were used, then the final left step was not used, so not all ingoing steps to  $a^{i-1}$  were used. But this implies that not all outgoing steps from  $a^{i-1}$  were used. Inductively this implies that not all incoming steps to  $a^0$  were used, a contradiction. So for each  $i > 0$ , all outgoing steps from  $a^i$  were used. Similarly, this is also true for  $i < 0$ . Hence every available step must have been used, so the path  $p$  is a base loop  $l'$  corresponding to the sequences  $c_1, \dots, c_m$  and  $d_1, \dots, d_n$ .  $\square$

Now using (5.1) and (5.2) it follows that for any two sequences  $c_1, \dots, c_m$  and  $d_1, \dots, d_n$ , with  $m, n \geq 1$ , the generating function for the corresponding loops  $l$  in  $G \wr \mathbb{Z}$  is equal to

$$t^{2c_1+2d_1} \binom{c_1+d_1}{c_1} P_{d_n} P_{c_m} P_{c_1+d_1+1} \prod_{i=1}^{m-1} t^{2c_{i+1}} \binom{c_i+c_{i+1}-1}{c_i-1} P_{c_i+c_{i+1}} \prod_{j=1}^{n-1} t^{2d_{j+1}} \binom{d_j+d_{j+1}-1}{d_j-1} P_{d_j+d_{j+1}}. \quad (5.3)$$

If  $m = 0$  and  $n \geq 1$ , the generating function is

$$t^{2d_1} P_{d_n} P_{d_1+1} \prod_{j=1}^{n-1} t^{2d_{j+1}} \binom{d_j+d_{j+1}-1}{d_j-1} P_{d_j+d_{j+1}}. \quad (5.4)$$

If  $m \geq 1$  and  $n = 0$  we get a similar generating function, and if  $m = n = 0$  we get  $P_1(t)$ . Now  $F(t)$  is the sum of these generating functions over all pairs of sequences  $c_1, \dots, c_m, d_1, \dots, d_n$

To calculate these we define some new power series  $\Omega_d(t)$  by

$$\Omega_d(t) = \sum P_{d_n}(t) \prod_{j=1}^{n-1} t^{2d_{j+1}} \binom{d_j+d_{j+1}-1}{d_j-1} P_{d_j+d_{j+1}}(t), \quad (5.5)$$

where the sum is over all sequences  $n, d_1, d_2, \dots, d_n$  with  $d_1 = d$ . Then it follows immediately from (5.3) and (5.4) that the generating function  $F(t)$  for loops in  $G \wr \mathbb{Z}$  is given by

$$F(t) = \left( \sum_{c,d=1}^{\infty} t^{2c+2d} \binom{c+d}{c} P_{c+d+1}(t) \Omega_d(t) \Omega_c(t) \right) + 2 \left( \sum_{d=1}^{\infty} t^{2d} P_d(t) \Omega_d(t) \right) + P_1(t). \quad (5.6)$$

So now we just need to calculate  $\Omega_d(t)$  for each positive integer  $d$ .

**Lemma 5.12.** *The series  $\Omega_d(t)$  satisfy the equations*

$$\Omega_d(t) = P_d(t) + \sum_{b=1}^{\infty} t^{2b} \binom{b+d-1}{d-1} P_{b+d}(t) \Omega_b(t). \quad (5.7)$$

*Proof.* We deduce this equation directly from the defining equation (5.5) of  $\Omega_d(t)$ . First, the contribution from the case where  $n = 1$  is  $P_{d_n} = P_{d_1} = P_d$ . The contribution from the case where  $n > 1$  and  $d_2 = b$  for some fixed positive integer  $b$  is

$$t^{2b} \binom{d+b-1}{d-1} P_{d+b}(t) \Omega_b(t).$$

Summing these contributions for all values of  $b$  yields the desired result.  $\square$

Using (5.7) we can calculate the coefficient of  $t^k$  in  $\Omega_d(t)$  in terms of coefficients of  $t^j$  in  $\Omega_b(t)$  where we only need to consider  $j, b$  satisfying  $2b+j \leq k$  (hence  $j \leq k-2$ ). This takes polynomial time using a simple recursive algorithm. Finally the coefficients of  $F(t)$  are given by (5.6).

## The Brin-Navas group $B$

In this section we adapt the previous algorithm to calculate the cogrowth series for the Brin-Navas group  $B$ . Again this is a polynomial time algorithm, however the polynomial has higher degree than the one for the algorithm of the previous section. Recall that the group  $B$  is defined as the semi-direct product

$$(\dots \wr \mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z} \wr \dots) \rtimes \mathbb{Z},$$

where  $s$  generates the copy of  $\mathbb{Z}$  on the right and the copies of  $\mathbb{Z}$  in the wreath product are generated by  $a_i = s^i a s^{-i}$  for  $i \in \mathbb{Z}$ , with larger values of  $i$  corresponding to copies of  $\mathbb{Z}$  which are further left in the wreath product.

We define the  $s$ -height of a word over the generating set  $\{a, s, a^{-1}, s^{-1}\}$  to be the sum of the powers of  $s$ . Before counting the total number of loops, we will count the number of upper loops, which we define below.

**Definition 5.13.** An *upper word* is a word in  $\{a, a^{-1}, s, s^{-1}\}^*$  which ends at  $s$ -height 0 in which any initial subword has non-negative  $s$ -height. An *upper loop* is an upper word which forms a loop in  $B$ . An *H-word* is an upper word which contains no  $a$  or  $a^{-1}$  steps at height 0. Finally, a *J-word* is an *H-word* which forms a loop in  $B$ .

Let  $G(t, y)$  be the generating function for upper loops, where  $t$  counts the total length and  $y$  counts the number of steps of the loop which end at height 0. For each positive integer  $n$ , let  $H_n(t, y)$  be the generating function for  $n$ -tuples  $w_1, w_2, \dots, w_n$  of  $H$ -words satisfying  $\overline{w_1 \dots w_n} = 1$ , where  $t$  counts the total length of  $w_1 \dots w_n$  and  $y$  counts the total number of steps which end at  $s$ -height 0. Given such a loop  $l$ , let the base loop  $l'$  be the subword consisting of all  $a$  and  $a^{-1}$  steps at  $s$ -height 0. Similarly to the previous algorithm, we let  $c_i$  be the number of steps in  $l'$  from  $a^{i-1}$  to  $a^i$ , and  $d_i$  be the number of steps in  $l'$  from  $a^{-i+1}$  to  $a^{-i}$ . Then the length  $|l'|$  of  $l'$  is equal to

$$\sum_{i=1}^m 2c_i + \sum_{j=1}^n 2d_j.$$

As in the previous subsection, for a given pair of sequences  $c_1, \dots, c_m, d_1, \dots, d_n$ , the number of corresponding base loops is equal to

$$\binom{c_1 + d_1}{c_1} \prod_{i=1}^{m-1} \binom{c_i + c_{i+1} - 1}{c_i - 1} \prod_{j=1}^{n-1} \binom{d_j + d_{j+1} - 1}{d_j - 1}. \quad (5.8)$$

Let  $l' = a_1 a_2 \dots a_{|l'|}$ , where each  $a_i \in \{a, a^{-1}\}$ , and let  $l = w_1 a_1 w_2 \dots a_{|l'|} w_{|l'|+1}$  be the decomposition where each step  $a_i$  is at  $t$ -height 0, so each word  $w_i$  is an  $H$ -word. We say that the  $a$ -height of one of the subwords  $w_i$  is equal to the integer  $p$  that satisfies  $a^p = a_1 \dots a_i$ . Then  $l$  is a loop if and only if for any height  $h$ , concatenating all of the words  $w_i$  at  $a$ -height  $h$  creates a loop. We define another generating function  $\Lambda_d(t, y)$  by

$$\Lambda_d(t, y) = \sum H_{d_n} \prod_{j=1}^{n-1} t^{2d_{j+1}} y^{2d_{j+1}} \binom{d_j + d_{j+1} - 1}{d_j - 1} H_{d_j + d_{j+1}}(t),$$

where the sum is over all sequences  $n, d_1, d_2, \dots, d_n$  with  $d_1 = d$ . In the same way as in the previous section we get the following equations. The corresponding equations in Section 5.3.1 are (5.6) and (5.7).

$$\begin{aligned} G(t, y) &= \sum_{c,d=1}^{\infty} t^{2c+2d} y^{2c+2d} \binom{c+d}{c} H_{c+d+1}(t, y) \Lambda_d(t, y) \Lambda_c(t, y) \\ &\quad + 2 \sum_{d=1}^{\infty} t^{2d} y^{2d} H_d(t, y) \Lambda_d(t, y) \\ &\quad + H_1(t, y). \end{aligned} \quad (5.9)$$

$$\Lambda_d(t) = H_d(t, y) + \sum_{b=1}^{\infty} t^{2b} y^{2b} \binom{b+d-1}{d-1} H_{b+d}(t, y) \Lambda_b(t). \quad (5.10)$$

Now to calculate  $G(t, y)$ , we just need to calculate the generating functions  $H_n(t, y)$ . For each  $k \in \mathbb{Z}_{\geq 0}$ , let  $J_k(t)$  be the generating function for  $J$ -loops in  $B$  which have exactly  $k$  steps which end at  $s$ -height 0. For each such word  $w$ , the number of ways of breaking it into  $n$   $H$ -words  $w_1, w_2, \dots, w_n$  such that  $w_1 \dots w_n = w$  is equal to

$$\binom{k+n-1}{n-1}.$$

Therefore, we can calculate each generating function  $H_n(t, y)$  in terms of the generating functions  $J_k(t)$  as follows:

$$H_n(t, y) = \sum_{k=0}^{\infty} y^k \binom{k+n-1}{n-1} J_k(t). \quad (5.11)$$

Finally, we will calculate the generating functions  $J_k(t)$ . Trivially we have  $J_0(t) = 1$ . For  $k > 0$ , let  $l$  be a loop counted by  $J_k(t)$ . Then  $l$  must contain exactly  $k$  steps which end at height 0, which are not  $a$  or  $a^{-1}$  steps. Hence they must all be  $s^{-1}$  steps. Therefore,  $l$  decomposes as

$$l = su_1 s^{-1} su_2 s^{-1} \dots su_k s^{-1},$$

where each word  $u_k$  ends at height 0 and never goes below height 0. Moreover, since  $l$  is a loop, we must have  $\overline{u_1 \dots u_k} = 1$ . Hence the word  $u = u_1 \dots u_k$  is counted by the generating function  $G(t, y)$ . Moreover, if  $u$  contains  $m$  steps which end at height 0, then there are exactly

$$\binom{m+k-1}{k-1}$$

ways to decompose  $u$  into subwords  $u_1, \dots, u_k$  which each end at height 0. Hence we get the equation

$$J_k(t) = \sum_{m=0}^{\infty} t^{2k} \binom{m+k-1}{k-1} [y^m] G(t, y). \quad (5.12)$$

Now using equations (5.9), (5.10), (5.11) and (5.12) as well as the base case  $J_0(t) = 1$ , we can calculate the coefficients of  $G(t, y)$  in polynomial time using a recursive algorithm. Finally we relate these coefficients to the total number of loops in  $B$  using the following lemma:

**Lemma 5.14.** *For each  $n$ , the number of loops  $b_n$  of length  $n$  in  $B$  over the generating set  $\{a, s, a^{-1}, s^{-1}\}$  is equal to*

$$b_n = \sum_{m=0}^{\infty} \frac{n}{m} [y^m][t^n] G(t, y). \quad (5.13)$$

*Proof.* Given a loop  $w = x_1 x_2 \dots x_n$ , let  $C_w$  be the set of cyclic permutations of  $w$

$$C_w = \{x_i \dots x_n x_1 \dots x_{i-1} \mid 1 \leq i \leq n\}.$$

We will show that for each  $w$ , the contribution to both sides of (5.13) from all loops  $w' \in C_w$  is the same. Let  $q$  be maximal such that  $x_1 \dots x_n = (x_1 \dots x_p)^q$  for some  $p$  (so  $pq = n$ ). Then  $|C_w| = p$ , so the loops in  $C_w$  contribute  $p$  to the left-hand side of (5.13). Now, if  $m$  is the number of times that this word  $w$  reaches its minimum  $t$ -height, then there are exactly  $m$  starting points  $x_i$  such that  $x_i \dots x_n x_1 \dots x_{i-1}$  is an upper word, and each of these contributes  $t^n y^m$  to  $G(t, y)$ . Hence there are  $m/q$  distinct cyclic permutations of  $w$  which are upper words, so together they contribute  $n/q = p$  to the right hand side of (5.13). Hence each set  $C_w$  contribute the same amount to both sides of (5.13). Since the sets  $C_w$  partition the set of loops in  $B$ , the equation (5.13) holds.  $\square$

Using (5.13) we can quickly calculate the coefficients of the cogrowth generating function  $C_B(t)$  using those of  $G(t, y)$ . In Table 5.1 we give the first 50 coefficients of this generating function. In fact we have 128 terms.

## A General Algorithm

Before we describe the algorithm which we use for Thompson's group  $F$ , we will describe a general algorithm which can be applied to any group admitting certain functions which can be computed very quickly. In the next subsection we will describe how we apply this algorithm to  $F$ . This algorithm could also be applied to any of the other groups which we have discussed, however for wreath products  $G \wr \mathbb{Z}$  and the Brin-Navas group  $B$ , it would be much less efficient than the specific algorithms described previously in this section.

Let  $G$  be a group with inverse closed generating set  $S$ . Let  $\Gamma(G, S)$  denote the Cayley graph of  $G$  with respect to the generating set  $S$ . We will often refer to this simply as  $\Gamma$ . We will assume that every loop has even length, however this algorithm could easily be altered to apply when this is not the case.

Our algorithm can be seen as a significantly more memory efficient version of the algorithm used by Haagerup, Haagerup and Ramirez-Solano [85]. Their algorithm is based on the following lemma:



---

1
4
28
232
2092
19864
195352
1970896
20275692
211825600
2240855128
23952786400
258287602744
2806152315048
30686462795856
337490492639512
3730522624066540
41422293291178872
461802091590831904
5167329622166765872
58012358366319158872
653272479274904359312
7376993667962247094112
83518163933592420945440
947797532286760923097848
10779770914124700529470264
122856228305621394118000520
1402877847412263986004347872
16048147989560391552043686160
183892883412730524613883088808
2110556326150834244975990231512
24259510831181186885644198829344
279244563297679787781517160899820
3218641495385722409923501191862264
37146337262307758446419466115479416
429227600058421313330040967935014416
4965493663308539362541734301378311648
57506535582014868288482236767840209688
666700108804771886996957763509359246064
7737176908622194648339548498436658811432
89878279784970230837678375953110478795352
1045033044367535197025078407316665177933928
12161645115366917947524997117208173413019632
141653302005285175865456465524239660635389712
1651274058730064356309776255817393993665780288
19264448513399180870635082273788105896265150480
224919270246185854430934219198103161122414157760
2627954546552385827255336138747466100454012242528
30726935577139566309665785537931570627782996384120
359517978960007312327796870699755173605904761839752

---

*Table 5.1:* The first 50 coefficients of the cogrowth series for the Brin-Navas group.

**Lemma 5.15.** *For each integer  $n \geq 0$  and each vertex  $v \in V(\Gamma)$ , let  $w_{n,v}$  be the number of walks of length  $n$  in  $\Gamma$  from  $e$  to  $v$ . Then the  $n$ th term  $c_n$  in the cogrowth series for  $G$  is given by*

$$c_n = w_{2n,e} = \sum_{v \in V(\Gamma)} w_{n,v}^2.$$

*Proof.* Given a loop  $\gamma = a_0 a_1 \dots a_{2n}$ , where each  $a_i \in V(\Gamma)$  and  $a_{2n} = a_0 = e$ , we define the midpoint of  $\gamma$  to be the vertex  $a_n$ . Then  $\gamma$  is made up of a walk of length  $n$  from  $e$  to its midpoint followed by a walk of length  $n$  from its midpoint to  $e$ . Hence, the number of loops in  $\Gamma$  of length  $2n$  with midpoint  $v$  is  $w_{n,v}^2$ . Therefore, the number  $c_n$  of walks of length  $2n$  in  $\Gamma$ , which start and end at  $e$  is given by

$$c_n = w_{2n,e} = \sum_{v \in V(\Gamma)} w_{n,v}^2.$$

□

Now we will describe the algorithm of Haagerup, Haagerup and Ramirez-Solano.

**HHR algorithm.** This algorithm inputs a positive integer  $n$  and outputs the number  $c_n$  of loops of length  $2n$  in  $\Gamma$ .

*Step 1.* Recursively calculate the number  $w_{n,v}$  of walks to each vertex  $v$  in  $B(e, n)$ , the ball of radius  $n$  in  $\Gamma$ .

*Step 2.* Return the sum the squares of the numbers  $w_{n,v}$ .

The fact that this returns the desired answer is due to Lemma 5.15.

Note also that for each walk from  $e$  to  $v$ , there is a corresponding walk from  $e$  to  $v^{-1}$ , so  $w_{n,v} = w_{n,v^{-1}}$ . Hence one can approximately halve the time and memory requirements of the algorithm by only calculating one of the values of  $w_{n,v}$  and  $w_{n,v^{-1}}$  for each  $v$ . Haagerup, Haagerup and Ramirez-Solano ran this algorithm on Thompson's group  $F$  for  $n \leq 24$ , thereby calculating the terms  $c_0, c_1, \dots, c_{24}$ .

The limiting factor of this algorithm is that it is necessary to store a large proportion of the ball of radius  $n$  in memory at the same time. As a result it is essentially impossible to get any more than 24 coefficients of the cogrowth series for Thompson's group  $F$  using this algorithm. On the other hand, for groups with polynomial growth, this algorithm takes polynomial time and space, so we have used this algorithm to compute 100 terms of the cogrowth sequence of the Heisenberg group  $H$ .

Our algorithm uses the same idea as the HHR algorithm however we only store the ball of radius  $k$  in memory, where  $k \approx n/2$ . Importantly, we do this without significantly increasing the running time of the program.

Let  $O$  be an object in the program which represents an element of  $G$ . Our algorithm requires the following functions to be implemented:

- $init()$ . This returns an object  $O$  which represents the identity in  $G$ .
- $val(O)$ . This returns a value which is uniquely determined by the element of  $G$  which the object  $O$  represents. In other words,  $val(O_1) = val(O_2)$  if and only if  $O_1$  and  $O_2$  represent the same element of  $G$ .
- For each generator  $\lambda \in S$ , we have an operation  $O.step_\lambda$ . If  $O$  initially represents the element  $g \in G$ , this changes  $O$  to an object which represents  $g\lambda$ .
- For each generator  $\lambda \in S$ , we have a function  $l_\lambda(O)$ , defined by  $l_\lambda(O) = |g\lambda| - |g|$ , where  $g$  is the element of  $G$  which  $O$  represents. That is,  $l_\lambda(O) = 1$  if applying  $\lambda$  moves  $g$  away from the identity.

The speed of our algorithm depends entirely on the efficiency of these functions and operations. For Thompson's group our implementations of each of these functions takes constant time on average. We will describe our implementations of these functions in Section 5.3.4. Note that we do not require an inverse of  $val$  to be implemented, which is important as we do not know how to implement such a function efficiently for Thompson's group.

It is important to carefully choose how to store an element  $g$  as an object  $O$  so that each function and operation can take constant time. If, for example, we choose to store  $g$  as an arbitrary list of generators  $\lambda_1, \lambda_2, \dots, \lambda_n$  which satisfy

$$\lambda_1 \lambda_2 \dots \lambda_n = g,$$

then  $step_\lambda$  can easily be implemented to run in constant time (it can just add  $\lambda$  to the end of the list), but the function  $val$  cannot, as it takes linear time to read through the list, which is necessary just to check if  $g$  is the identity.

The memory requirements of our algorithm do not depend significantly on our implementations of the functions above, as more memory is used to store the numbers such as the number of paths to each vertex. Hence it is most important to minimise the time requirements of the functions even if that requires using more memory.

We are now ready to describe the first steps in our algorithm.

**Numloops algorithm.** This algorithm inputs a positive integer  $n$ , and outputs the number  $c_n$  of loops of length  $2n$  in  $\Gamma$ . This algorithm relies on the functions  $init$ ,  $val$ ,  $step_\lambda$  and  $l_\lambda$ .

- Step 1.* Assign an arbitrary order to the generating set  $S$  and set  $k = \lceil \frac{n}{2} \rceil$ .
- Step 2.* Using a simple recursive algorithm, construct an associative array  $A_{n-k}$ , implemented as a hash table, with a key value pair  $(k_g, a_g)$  for each element  $g \in G$  within the ball of radius  $n - k$ . The key  $k_g$  is given by  $val(O)$  where  $O$  is any object which represents  $g$  and the value  $a_g$  is equal to the number of walks of length  $n - k$  in  $\Gamma$  from  $e$  to  $g$ . We will write  $a_g = A[k_g]$ . For a number  $x$  which is not a key in  $A_{n-k}$ , we set  $A[x] = 0$ .
- Step 3.* Construct a tree  $T_k$  which contains one vertex  $v_g$  for each element  $g$  of  $G$  within the ball of radius  $k$ , such that each vertex  $v_g$ , apart from  $v_e$ , is connected to exactly one vertex  $v_h$  satisfying  $|h| = |g| - 1$  and  $h^{-1}g \in S$ . If there are multiple possible choices of  $h$ , we choose the element  $h$  which minimises  $\lambda = h^{-1}g$ , according to the order we assigned in step 1. The edge  $(h, g)$  is then labelled with  $\lambda$ . Each vertex  $v_g$  is also labelled with the number  $p(v_g)$  of paths of length  $k$  in  $\Gamma$  from  $e$  to  $g$ .

Before we describe the following steps in the algorithm, we will make some comments on these functions and prove a lemma required for the next step.

Note that steps 2 and 3 create structures which contain essentially the same data, as for each element  $g \in G$  satisfying  $|g| \leq n - k$ , the values  $p(v_g)$  and  $A[k_g]$  are equal. However, the use of different data structures is important for the following steps in the algorithm.

Step 4 of the algorithm relies on the following lemma:

**Lemma 5.16.** *For each element  $g \in G$ , the number of paths of length  $n$  in  $\Gamma$  from  $e$  and  $g$  is equal to*

$$\sum_{h \in G} p(v_h) A[k_{gh}]. \quad (5.14)$$

*Proof.* Each path of length  $n$  from  $e$  to  $g^{-1}$  in  $\Gamma$  can be written in a unique way as a path of length  $k$  from  $e$  to some vertex  $h$  in  $\Gamma$  followed by a path of length  $n - k$  from  $h$  to  $g^{-1}$ . For a given  $h$ , the number of these paths is equal to  $p(v_h) A[k_{h^{-1}g^{-1}}] = p(v_h) A[k_{gh}]$ . Summing over all vertices  $h \in \Gamma$  yields the desired result.  $\square$

Note also that the summand in (5.14) is 0 unless  $|h| \leq k$  and  $|gh| \leq n - k$ , so we only need to sum over values of  $h$  which satisfy these two inequalities. We are now ready to describe the last steps of our algorithm.

**Numloops algorithm (steps 4 and 5).**

*Step 4.* In this step we create a function  $numpaths(O, d)$  whose input is an object  $O$  and a positive integer  $d$ , which, assuming that  $d = |g|$ , outputs the number of paths of length  $n$  in  $\Gamma$  from  $e$  to  $g$ , where  $g$  is the group element represented by  $O$ . During the calculation of  $numpaths$  the object  $O$  may change, but at the end it must represent the same group element  $g$ . Lemma 5.16 implies that this is equivalent to calculating (5.14). To calculate this expression, we perform a depth first search of the tree  $T_k$ , skipping any sections where we can be sure that there are no vertices  $v_h$  such that  $h$  satisfies the two inequalities  $|h| \leq k$  and  $|gh| \leq n - k$ . We start the search at the root vertex  $v_e$  of  $T_k$  and initialise  $r = 0$  and  $total = 0$ . Whenever we move from a vertex  $v_h$  to  $v_{h\lambda}$  we change  $d$  to  $d + l_\lambda(O)$  and then apply the operation  $O.step_\lambda$ . That way whenever we are at a vertex  $v_h$ , the object  $O$  represents  $gh$  and  $d = |gh|$ . We also increase  $x$  by 1 whenever we move to a child vertex and decrease  $x$  by 1 when we backtrack so that we always have  $x = |h|$ . Then we add  $p(v_h)A[k_{gh}] = p(v_h)A[val(O)]$  to the sum  $total$  if and only if  $d \leq n - k$ , since  $x = |h| \leq k$  for every vertex  $v_h$  in  $T_k$ . Since  $d$  decreases by at most 1 when we move to a child vertex, and  $x$  always increases by 1, the value  $x + d$  never decreases when we move to a child vertex. So if  $x + d > n$  when we are at a vertex  $v_h$ , then we do not traverse the children of  $v_h$ . At the end of the search we return to the root vertex and then return the value  $total$ .

*Step 5.* For the last step we just need to add up the value of  $numpaths$  for every vertex  $g$  in the ball of radius  $n$  such that  $|g|$  has the same parity as  $n$ . To accomplish this we perform a depth first search of the tree  $T_n$ , which is defined in the same way as  $T_k$ . However, we do not explicitly construct  $T_n$  as doing so would use too much memory. In order to perform the depth first search, we just need a function  $isedge_\lambda(O)$  for each  $\lambda \in S$  which returns 1 if and only if there is an outward edge from  $v_g$  to  $v_{g\lambda}$  in  $T_n$ , where  $g$  is the group element that  $O$  represents. This will be the case if and only if  $|g\lambda| = |g| + 1$  and  $|g\lambda\mu| = |g\lambda| + 1$  for each  $\mu \in S$  with  $\mu < \lambda^{-1}$ . We test this using the functions  $l_\lambda$ ,  $step_\lambda$  and  $l_\mu$ . During the depth first search, we keep track of the distance  $d = |g|$ , where  $g$  is the group element represented by  $O$ . Now, to calculate the number  $numloops$  of loops of length  $2n$ , we first set  $numloops = 0$ , then run the depth first search, and when we visit each vertex of  $T_n$ , add  $numpaths(O, d)^2$  to  $numloops$ . At the end of this process  $numloops$  is equal to the number of loops of length  $2n$ , so we return  $numloops$  and terminate the algorithm.

The advantage of this algorithm is that it only stores  $T_k$  and  $A_{n-k}$  in

memory, rather than all of  $T_n$ . This also allows us to parallelise step 5, constructing  $T_k$  and  $A_{n-k}$  separately for each core.

## Thompson's group $F$

In this section we describe how the object  $O$ , the operation  $step_\lambda$  and the functions  $val$  and  $l_\lambda$  are implemented for Thompson's group  $F$ . We use the standard generating set  $S = \{a, b, a^{-1}, b^{-1}\}$ , which yields the presentation

$$F = \langle a, b \mid a^2ba^{-2} = baba^{-1}b^{-1}, a^3ba^{-3} = ba^2ba^{-2}b^{-1} \rangle.$$

Recall the forest representation of elements of  $F$  given by Belk and Brown [19] and described in Section 5.2.

**Definition 5.17.** A *forest diagram* is a pair  $(\mathcal{F}_1, \mathcal{F}_2)$  of sequences of planar binary trees such that each sequence has the same total number of leaves, with one tree highlighted in each sequence.

In order to move between adjacent elements of Thompson's group, we need to understand the effect the generators  $a$ ,  $b$ ,  $a^{-1}$  and  $b^{-1}$  have on the forest diagram. To describe this, let  $(\mathcal{F}_1, \mathcal{F}_2)$  be a forest digram, let  $T_j$  be the highlighted tree in  $\mathcal{F}_1$ , and let  $T_{j-1}$  and  $T_{j+1}$  be the adjacent trees. The element  $a$  (resp.  $a^{-1}$ ) unhighlights  $T_j$  and highlights  $T_{j+1}$  (resp.  $T_{j-1}$ ) instead. The element  $b$  adds a new vertex  $v$  to  $\mathcal{F}_1$ , making  $T_j$  its left subtree and  $T_{j+1}$  its right subtree, then highlights the new tree formed. Conversely,  $b^{-1}$  deletes the root vertex of  $T_j$  forming two new trees, and highlights the one on the left. In the case that  $T_j$  is a single vertex, the second equivalence operation is used on this vertex before  $b^{-1}$  is applied.

Now we are ready to describe how we have implemented the object  $O$ , the operation  $step_\lambda$  and the functions  $val$  and  $l_\lambda$  for Thompson's group  $F$ . For  $O$  we simultaneously store an element  $g \in F$  as a graph  $P$ , a pair of binary strings  $s_1, s_2$  and two integers  $p_1$  and  $p_2$ . The vertices of  $P$  are the vertices of the trees in  $\mathcal{F}_1$  and  $\mathcal{F}_2$ , where  $(\mathcal{F}_1, \mathcal{F}_2)$  is the reduced forest diagram corresponding to  $g$ . The edges of  $P$  fall into three types: one type of edge in  $P$  corresponds to the edges of the trees, another type of edge joins matching leaves and a third type of edge joins the root vertex of each tree to the root vertex of each adjacent tree in the sequence. Now we define the binary strings  $s_1$  and  $s_2$ . A single binary tree with  $m$  leaves corresponds to a unique binary string  $s$  of length  $2m - 2$  with the property that  $s$  has an equal number of 1's and 0's and the number of 1's in any initial substring is at least equal to the number of 0's in that substring. This is defined by doing a depth first search of the tree and writing a 1 whenever we move down an

edge from a vertex to its left subtree and writing a 0 whenever we backtrack along such an edge. To convert a sequence of binary trees to a binary string, we first convert each individual tree to a binary string, insert the string 01 before each such string, then concatenate the results. We then change the 01 before the string corresponding to the highlighted tree to 00.  $s_1$  and  $s_2$  are the binary strings corresponding to the sequences  $\mathcal{F}_1$  and  $\mathcal{F}_2$ . We also store the numbers  $p_1$  and  $p_2$  in  $O$ , which define the positions of the 00 before the highlighted tree in each of  $s_1$  and  $s_2$ . The strings  $s_1$  and  $s_2$  each have length at most  $2n$ , so they can be represented as 64 bit numbers as long as  $n \leq 32$ .

The operation  $step_\lambda$  is defined easily for the effect on the graph  $P$ . The effect on the binary strings  $s_1$  and  $s_2$  is a bit more complicated and requires some bit shifting. The entire length of an element of Thompson's group  $F$  can be determined by its forest diagram, as shown in [19], so we could use this to determine  $l_\lambda$  by using the graph  $P$  and simply subtracting the calculated length  $|g|$  from the length we calculate for  $|g\lambda|$ . In fact we do it more efficiently than this, as the difference  $|g\lambda| - |g|$  is determined entirely by the highlighted tree and the surrounding trees. Finally,  $val(O)$  simply returns the pair  $(s_1, s_2)$ .

We have used this algorithm to compute the first 32 terms of the cogrowth sequence for Thompson's group  $F$ . These coefficients are shown in Table 5.2. This is 7 further terms than given in [128].

## Possible cogrowth of the groups $F$ and $B$

In this section we will show that if  $a_0, a_1, \dots$  is the cogrowth sequence for Thompson's group  $F$ , then for any real numbers  $a < 1$  and  $\lambda > 1$ , the inequality

$$a_n < 16^n \lambda^{-n^a}$$

holds for all sufficiently large integers  $n$ . As a result, if Thompson's group is amenable, then the sequence cannot grow at the rate

$$16^n \lambda^{-n^a},$$

For any fixed  $a < 1$ . We then show that the same result holds for the Brin-Navas group  $B$ , which is known to be amenable. This result follows quite readily from results in [170] and [169], however we will need some definitions before we can see how they apply.

Let  $G$  be a group with finite generating set  $S$ . Then we define the function  $\phi_S : \mathbb{Z}_{>0} \rightarrow \mathbb{R}_{>0}$  by setting  $\phi_S(n)$  to be the probability that a random walk

Coefficients
1
4
28
232
2092
19884
196096
1988452
20612364
217561120
2331456068
25311956784
277937245744
3082543843552
34493827011868
389093033592912
4420986174041164
50566377945667804
581894842848487960
6733830314028209908
78331435477025276852
915607264080561034564
10750847942401254987096
126768974481834814357308
1500753741925909645997904
17833339046478612301547884
212663448005862463186139032
2544535423071442709522261116
30542557512715560857221200908
367718694478039302564802454628
4439941127401928226610731571976
53756708216952135677787623701460

*Table 5.2:* Terms in the cogrowth sequence of Thompson's group  $F$ .

in  $(G, S)$  of length  $2n$  finishes at the origin. In other words,  $|S|^{2n}\phi_S(n)$  is the number of loops of length  $2n$  in the Cayley graph  $\Gamma(G, S)$ .

Given two (non-increasing) functions  $\phi_1$  and  $\phi_2$ , we say that  $\phi_1 \preceq \phi_2$ , if there is some  $C \in \mathbb{R}_{>0}$  such that  $\phi_1(n) \leq C\phi_2(n/C)$ , where each  $\phi_i$  is



extended to the reals by linear interpolation. Finally we say that  $\phi_1 \approx \phi_2$  if both  $\phi_1 \preceq \phi_2$  and  $\phi_2 \preceq \phi_1$ . We recall Theorem 1.3 from [169]:

**Theorem 5.18.** *Let  $G$  be a group with finite, symmetric generating set  $S$ , let  $H$  be a subgroup of  $G$  and let  $T$  be a finite symmetric generating set of  $H$ . Then*

$$\phi_S \preceq \phi_T.$$

Note that if  $S$  and  $T$  are finite generating sets for the same group  $H = G$ , then it follows that  $\phi_T \approx \phi_S$ .

The other result we need concerns wreath products with  $\mathbb{Z}$ . In [170], Pittet and Saloff-Coste show (in a remark just below Theorem 8.11) that for a finite generating set  $T$  of  $\mathbb{Z} \wr_d \mathbb{Z}$ , we have

$$\phi_T(n) \approx \exp \left( -n^{d/(d+2)} (\log n)^{2/(d+2)} \right).$$

Now, since  $\mathbb{Z} \wr_d \mathbb{Z}$  is a subgroup of Thompson's group  $F$ , we must have

$$\phi_S(n) \preceq \phi_T(n) \approx \exp \left( -n^{d/(d+2)} (\log n)^{2/(d+2)} \right),$$

where  $S$  is the standard generating set of  $F$ . Hence, for any positive integer  $d$ , there is a positive real number  $C$  such that

$$\phi_S(n) \leq C \exp \left( -(n/C)^{d/(d+2)} (\log(n/C))^{2/(d+2)} \right).$$

Now we are ready to prove our theorem.

**Theorem 5.19.** *Let  $a_n$  be the number of loops of length  $2n$  in the standard Cayley graph for Thompson's group. Then for any real numbers  $a < 1$  and  $\lambda > 1$ , the inequality*

$$a_n < 16^n \lambda^{-n^a}$$

*holds for all sufficiently large integers  $n$ .*

*Proof.* Let  $d$  be a positive integer such that  $\frac{d}{d+2} > a$ . Then there is some  $C \in \mathbb{R}_{>0}$  such that

$$\phi_S(n) \leq C \exp \left( -(n/C)^{d/(d+2)} \log(n/C)^{2/(d+2)} \right)$$

for all  $n \in \mathbb{Z}_{>0}$ . For  $n$  sufficiently large, we have  $\log(n/C) > 0$ , so

$$\begin{aligned} C \exp \left( -(n/C)^{d/(d+2)} \log(n/C)^{2/(d+2)} \right) &< C \exp \left( -(n/C)^{d/(d+2)} \right) \\ &= \exp \left( \log(C) - C^{-d/(d+2)} n^{d/(d+2)} \right), \end{aligned}$$

Hence, for all  $n$  sufficiently large we have

$$\phi_S(n) < \exp(-n^\alpha).$$

Therefore,

$$a_n = 16^n \phi_S(n) < 16^n \exp(-n^\alpha)$$

□

Note that the same result holds if we replace Thompson's group  $F$  with the Brin-Navas group  $B$ , since it also contains every wreath product  $\mathbb{Z} \wr_d \mathbb{Z}$  as a subgroup. For Thompson's group  $F$ , Theorem 5.19 is immediately true if  $F$  is non-amenable, as then its cogrowth sequence has growth rate less than 16. If  $F$  is amenable, then Theorem 5.19 implies that its cogrowth sequence grows remarkably slowly for a sequence with exponential growth rate 16. Since  $B$  is amenable, its cogrowth sequence must have growth rate 16, so Theorem 5.19 restricts the possible sub-exponential behaviour of this cogrowth sequence.

## Empirical analysis using differential approximations

In this section we make a numerical study of the cogrowth sequences for the Brin-Navas group  $B$  and Thompson's group  $F$  using the series we computed with algorithms described in Section 5.3. A more detailed numerical analysis of these sequences is given in [92], as is a similar analysis of the other cogrowth sequences considered in this chapter.

### Analysis of the Brin-Navas group $B$ .

This is an amenable group introduced independently by Navas [161] and Brin [36], so we call it the Brin-Navas group  $B$ . It has 2 generators, so the growth rate of its cogrowth sequence is 16. We gave a polynomial-time algorithm to generate terms of this sequence in Section 5.5.1, and have used this to generate 128 terms of the cogrowth sequence. We then used the method of series extension, described in Section 2.5.4 and [124], to estimate a further 590 ratios, the last of which we expect to be accurate to 1 part in  $5 \times 10^7$ , while all earlier ratios have a lower associated error. We denote the  $n$ th ratio by  $r_n$ . We first show a plot of the modified ratios

$$r_n^{(1)} = n \cdot r_n - (n-1) \cdot r_{n-1}. \quad (5.15)$$

against  $1/n$  in Figure 5.1. Even if we knew nothing about the asymptotics of this group, the curvature of this plot provides strong evidence for a sub-exponential term, and we have proved that it cannot be a regular stretched-exponential term.

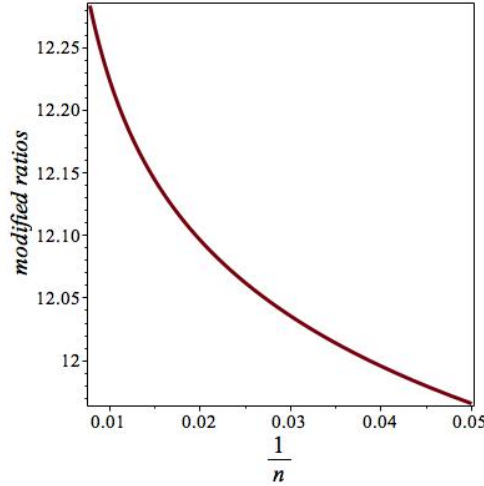


Figure 5.1: The first 128 modified ratios of the Brin-Navas group  $B$  vs.  $1/n$ .

That is to say, the asymptotics for this series must grow more slowly than

$$c_n \sim c \cdot \mu^n \cdot \kappa^{n^\sigma} \cdot n^g,$$

where  $\mu = 16$ ,  $0 < \sigma < 1$ , and  $0 < \kappa < 1$ . Possible behaviour might be

$$c_n \sim c \cdot \mu^n \cdot \kappa^{n/\log n} \cdot n^g,$$

corresponding to a *numerical* value  $\sigma = 1$ , which of course hides the logarithmic component.

In that case the ratios will be

$$r_n = \frac{c_n}{c_{n-1}} \sim \mu \left( 1 + \frac{\text{constant}}{\log n} + \frac{g}{n} + \dots \right).$$

Note that we do not insist that the first correction term is  $O(1/\log n)$ , it could be a power of a logarithm, or some other weakly decreasing function, but it cannot have a power-law increase. For our purposes it suffices to take this term to be  $O(1/\log n)$ . We show the modified ratios (this gets rid of the  $O(1/n)$  term in the asymptotics) in Figures 5.2 and 5.3 which are the same plot, but the first uses only the 128 exact coefficients, while the second uses the exact plus predicted ratios. From the first plot, it is clear that it would

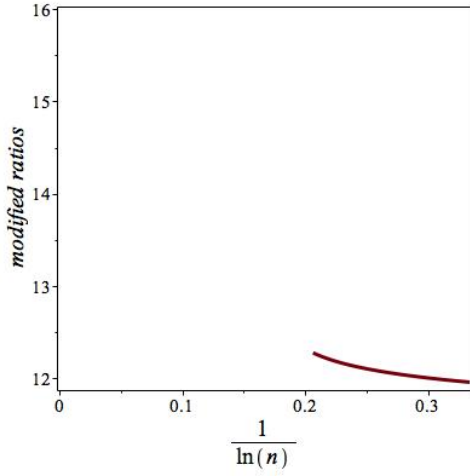


Figure 5.2: The first 128 modified ratios for the Brin-Navas group  $B$  vs.  $1/\log n$ .

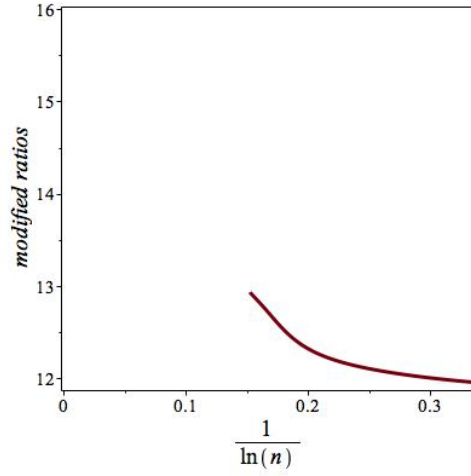


Figure 5.3: The first 718 modified ratios for the Brin-Navas group  $B$  vs.  $1/\log n$ .

be an article of faith that the locus is going to 16 as  $n \rightarrow \infty$ . By contrast, the second plot makes this conclusion far more plausible.

We next try to estimate the exponent  $\sigma$ , which should be 1, as follows: Assume that the coefficients  $c_n$  take the asymptotic form

$$c_n \sim c \cdot \mu^n \cdot \kappa^{n^\sigma \log(n)^\delta} \cdot n^g,$$

where  $c, \mu, \kappa, \sigma, \delta, g$  are positive real constants with  $\sigma \leq 1$ . Forming the ratio of ratios

$$rr_n^{(1)} = \frac{r_n}{r_{n-1}} = 1 + \frac{\log \kappa \cdot \log^\delta n}{n^{2-\sigma}} \left( \sigma(\sigma-1) + \frac{\delta(2\sigma-1)}{\log n} + \frac{\delta(\delta-1)}{\log^2 n} \right) - \frac{g}{n^2} + o(1/n^2) \quad (5.16)$$

eliminates the constants  $c$  and  $\mu$ . If we now form the sequence

$$t_n = \frac{rr_n^{(1)} - 1}{rr_{n-1}^{(1)} - 1} \quad (5.17)$$

this eliminates the base  $\kappa$  of the stretched-exponential term, and in fact

$$n(t_n - 1) \sim \sigma - 2 + \frac{\delta}{n \log n}.$$

So plotting  $n(t_n - 1)$  against  $1/(n \log n)$  should give an estimate of  $\sigma - 2$ .

With the 128 known terms, the estimators of  $\sigma - 2$  are shown in Figure 5.4 and show no evidence of approaching the expected value of  $-1$ . If however

we use twice as many terms, so using the next 128 predicted ratios, we get the plot shown in Figure 5.5, which *is* plausibly approaching  $-1$ .

This highlights the value of numerically predicting further terms wherever possible.

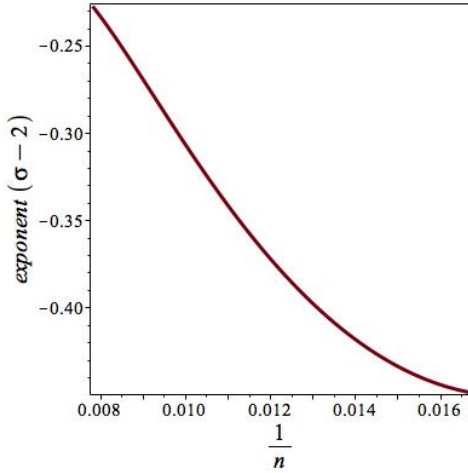


Figure 5.4: Estimates of  $\sigma - 2$  from 128 terms of the Brin-Navas group  $B$ .

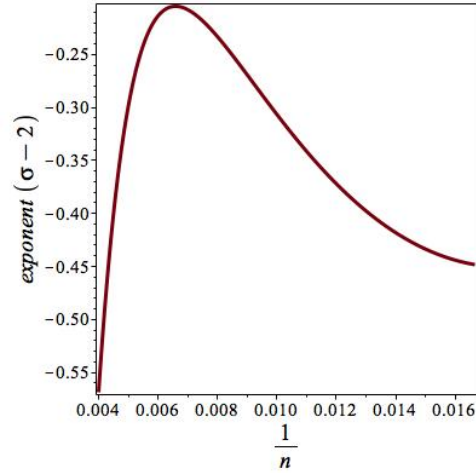


Figure 5.5: Estimates of  $\sigma - 2$  from 256 terms of the Brin-Navas group  $B$ .

## Analysis of Thompson's group $F$

For Thompson's group  $F$  it is known that the series grows exponentially like  $\mu^n$ . If  $\mu = 16$ , the group is amenable. As we proved in Theorem 5.19, if Thompson's group is amenable, there cannot be a sub-dominant term of the form  $\kappa^{n^\sigma}$  with  $0 < \sigma < 1$ .

We first study the modified ratios, defined by (5.15). The modified ratio plot against  $1/n$  is shown in Figure 5.6 and displays considerable curvature. By contrast, the same data plotted against  $n^{-1/5}$ , and shown in Figure 5.7 shows curvature in the opposite direction. This suggests that the plot should be (asymptotically) linear when plotted against  $1/n^{1-\sigma}$  for  $0 < \sigma < 4/5$ . This is strong evidence for the presence of a conventional stretched-exponential term  $\kappa^{n^\sigma}$  or perhaps  $\kappa^{n^\sigma \log(n)^\delta}$  where  $0 < \sigma < 4/5$ . As mentioned above, the presence of such a term is incompatible with amenability. This is our first piece of evidence that the group is not amenable. Note too that this is quite different to the behaviour observed for the coefficients of the Brin-Navas group  $B$ .

In our subsequent analysis, we use both the exact coefficients and the extrapolated coefficients. While all extrapolated terms can be used in cal-

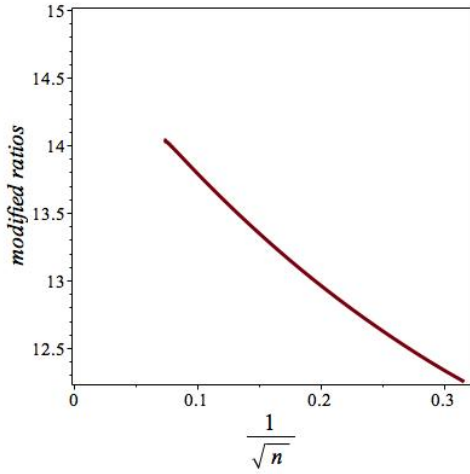


Figure 5.6: Modified ratios vs.  $1/n$  for Thompson's group  $F$ .

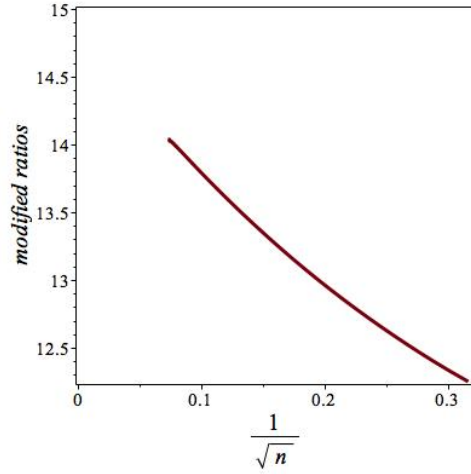


Figure 5.7: Modified ratios vs.  $n^{-1/5}$  for Thompson's group  $F$ .

culating the ratios, once one calculates first and second differences, errors are amplified, and so fewer terms can be used. That is why we quote the number of terms used for different calculations, as it is only to the quoted order that we are confident that the calculated quantities are accurate to graphical accuracy.

To estimate the exponents in the stretched-exponential term we use the procedure described by eqn. (5.17) and the discussion below. This procedure allows for the presence of a confluent power of a logarithm, so that the stretched-exponential term is  $\kappa^{n^\sigma \log^\delta n}$ . In this way, based on a series of length 80, we show a plot of estimators of  $2-\sigma$  against  $1/n$  in Figure 5.8. To estimate  $\delta$ , we form the sequence

$$n \log^2 n(n(t_n - 1) - (n - 1)(t_{n-1} - 1)) \sim -\delta + O(1/\log n).$$

These estimators of  $-\delta$  are plotted against  $1/n$  in 5.9. Extrapolating these, we estimate  $\sigma \approx 1/2$ , and  $\delta \approx 1/2$ . Recall that this is exactly the stretched-exponential behaviour of  $(\mathbb{Z} \wr \mathbb{Z}) \wr \mathbb{Z}$ .

Reverting to the modified ratios, briefly discussed above, we plot these against  $1/\sqrt{n}$  in Figure 5.10, using 186 terms. One observes that the plot still displays a little curvature, but in Figure 5.11 the plot of these same modified ratios against  $\sqrt{\log n/n}$ , is essentially linear. This is the appropriate power to extrapolate against, given our estimates of the stretched-exponential exponents. Extrapolating this to  $n \rightarrow \infty$  we estimate the limit, which gives the growth constant, to be  $14.8 - 15.1$ . This is well away from 16, which would be required for amenability.

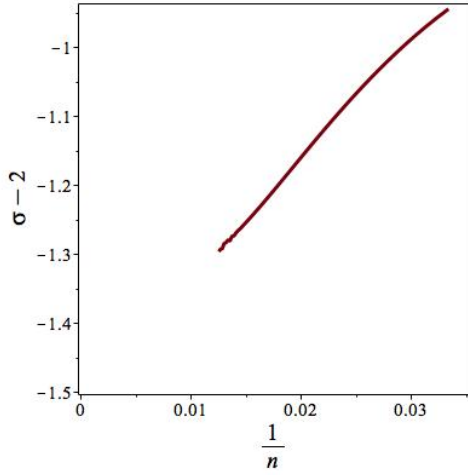


Figure 5.8: Estimators of  $\sigma - 2$  for Thompson's group  $F$  vs.  $1/n$ .

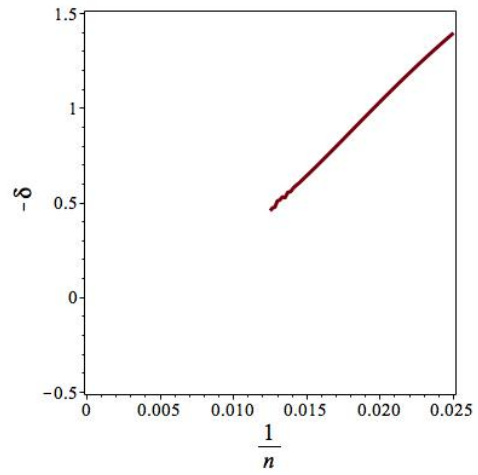


Figure 5.9: Estimators of  $-\delta$  for Thompson's group  $F$  vs.  $1/n$ .

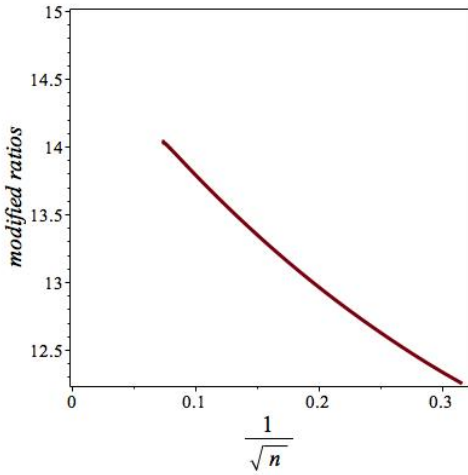


Figure 5.10: The first 186 modified ratios for Thompson's group  $F$  plotted against  $1/\sqrt{n}$ .

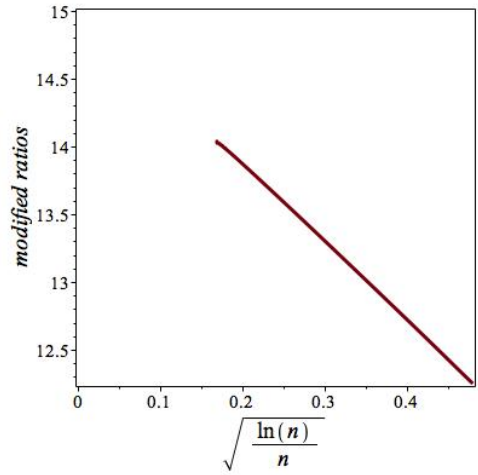


Figure 5.11: The first 186 modified ratios for Thompson's group  $F$  plotted against  $\sqrt{\log n/n}$ .

One simple test for amenability uses the fact that the ratio of successive coefficients asymptotes to the growth constant  $\mu$ . For the lamplighter group, this ratio behaves as

$$r_n^{(L)} = 9 \left( 1 + \frac{c}{n^{2/3}} + o\left(\frac{1}{n^{2/3}}\right) \right).$$

For  $\mathbb{Z} \wr \mathbb{Z}$  one has

$$r_n^{(2)} = 16 \left( 1 + \frac{c \cdot \log^{2/3} n}{n^{2/3}} + o \left( \frac{\log^{2/3} n}{n^{2/3}} \right) \right),$$

and for the triple wreath product,  $\mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z}$  the corresponding result is

$$r_n^{(3)} = 36 \left( 1 + \frac{c \sqrt{\log n}}{n^{1/2}} + o \left( \frac{\sqrt{\log n}}{n^{1/2}} \right) \right),$$

while for Thompson's group  $F$  essentially all we know is

$$r_n = \mu (1 + \text{lower order terms}),$$

though from the above analysis there are hints that the correction term is similar to that of the triple wreath product of  $\mathbb{Z}$ .

So, a simple test for amenability is to look at the three quotients

$$\frac{9r_n}{16r_n^{(L)}}, \quad \frac{r_n}{r_n^{(2)}}, \quad \text{and} \quad \frac{9r_n}{4r_n^{(3)}}.$$

If Thompson's group  $F$  is amenable, these quotients should all go to 1. In Figures 5.12, 5.13, 5.14 we show these ratios plotted against  $\sqrt{\log n/n}$ , which is the appropriate power, though this choice is not critical. The ratios do not appear to be going to 1 in any of the three cases. For all cases we have used 200 ratios. To do this, we used the extended ratios for Thompson's group  $F$  and also extended the ratios for  $\mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z}$  from the known 132 ratios. Indeed, all three cases are consistent with a limit around  $0.93 \pm 0.02$ , corresponding to  $\mu = 14.9 \pm 0.3$ . This is entirely consistent with our previous estimate of  $\mu \approx 15.0$ .

## Conclusion

We have given polynomial-time algorithms to generate terms of the cogrowth series for several groups. In particular, we have given the first series for the Brin-Navas group  $B$ . We have also given an improved algorithm for the coefficients of Thompson's group  $F$ , giving 32 terms of the cogrowth series, extending previous enumerations by 7 terms. For Thompson's group  $F$  we proved that, if the group is amenable, there cannot be a sub-dominant stretched exponential term in the asymptotics. The numerical data however provides compelling evidence for the presence of such a term. This observation suggests a potential path to a proof of non-amenability.



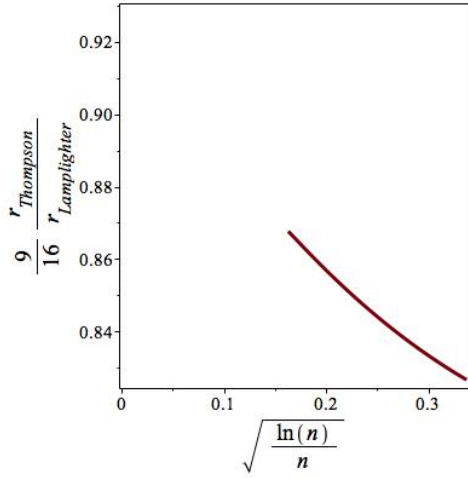


Figure 5.12: Quotient of  $F$  and  $L$  ratios using 200 terms.

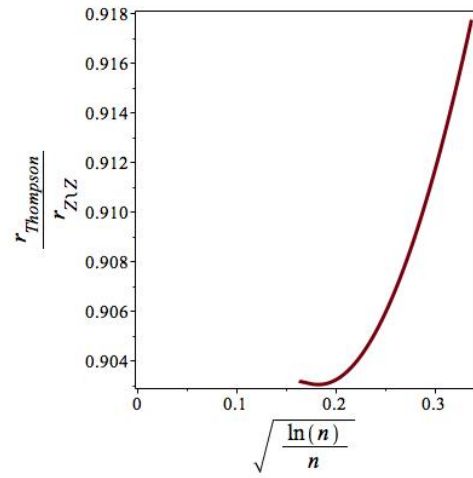


Figure 5.13: Quotient of  $F$  and  $\mathbb{Z} \wr \mathbb{Z}$  ratios using 200 terms.

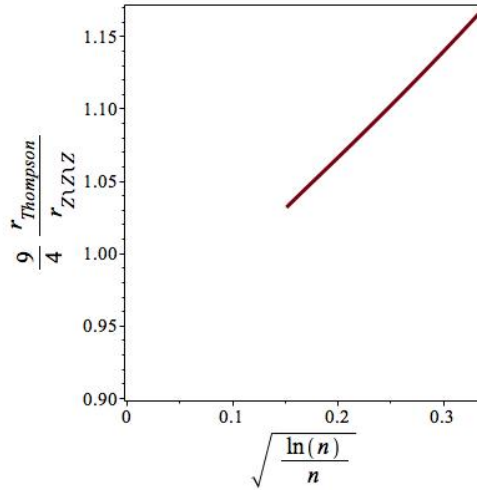


Figure 5.14: Quotient of  $F$  and  $(\mathbb{Z} \wr \mathbb{Z}) \wr \mathbb{Z}$  ratios using 200 terms.

We have extended the sequence of 32 terms for group  $F$  by a further 200 terms (or, as appropriate, 200 ratios of successive terms), which we demonstrate are sufficiently accurate for the graphical approaches to analysis that we have taken.

A numerical study of the cogrowth sequence  $c_n$  leads us to suggest that

$$c_n \sim c \cdot \mu^n \cdot \kappa^{n^\sigma \log^\delta n} \cdot n^g,$$

where  $\mu \approx 15$ ,  $\kappa \approx 1/e$ ,  $\sigma \approx 1/2$ ,  $\delta \approx 1/2$ , and  $g \approx -1$ . The growth constant  $\mu$  must be 16 for amenability. These two approaches to the study of

amenability lead us to conjecture that Thompson's group  $F$  is not amenable.

The difficulties we encountered in analysing the Brin-Navas group  $B$  demonstrate that there exist groups whose cogrowth series are difficult to analyse. Nevertheless, through our empirical analysis, we were able to extract asymptotic properties of this sequence which highlight these difficulties. Furthermore, the cogrowth series for Thompson's group  $F$  did not behave like that of  $B$ , and indeed appeared to have a stretched exponential term with exponent values that were readily estimable. While we cannot rule out the presence of some previously unsuspected pathology in the asymptotic form, we believe that we have presented strong evidence for the conjecture that Thompson's group  $F$  is not amenable.

# Chapter 6

## Combinatorial Stieltjes moment sequences

### Introduction to Stieltjes moment sequences

Given a combinatorial class which is enumerated by some sequence  $\mathbf{a} = (a_n)_{n \geq 0}$ , one of the first questions that one may ask is whether the growth rate

$$\mu = \lim_{n \rightarrow \infty} \sqrt[n]{a_n}$$

exists. In many cases the existence of this limit can be proved using the super-multiplicative property  $a_{m+n} \geq a_m a_n / a_0$ , along with Fekete's lemma, as long as the sequence is also known to grow at most exponentially. For example, this method is used for principal permutation classes [154] and various classes of maps [42]. Super-multiplicativity of the sequence implies that the numbers  $\sqrt[n]{a_n}$  all lie strictly below the limit  $\mu$ , so this immediately yields a method to compute arbitrarily strong lower bounds for  $\mu$  given arbitrarily many terms of the sequence. Unfortunately these bounds tend to converge very slowly. For example, consider the sequence  $c_0, c_1, \dots$  of Catalan numbers. The value

$$\sqrt[100]{c_{100}} \approx 3.7113$$

is still a long way from the exact growth rate  $\mu = 4$ .

In many super-multiplicative combinatorial sequences, we empirically observe the stronger property that the sequence is log convex, meaning that the sequence of ratios  $r_n = a_n / a_{n-1}$  is non-decreasing. If this is the case it immediately implies that the growth rate  $\mu$  exists (though it may be infinite) and that it is equal to the limit  $\lim_{n \rightarrow \infty} a_n / a_{n-1}$ . In general these ratios converge much more quickly than the values  $\sqrt[n]{a_n}$ , which is useful both for approximating the growth constant as well as computing lower bounds for  $\mu$ . Taking

the example of Catalan numbers again, the value  $c_{100}/c_{99} \approx 3.9406$ , which is significantly closer to the exact growth rate 4 than  $\sqrt[100]{c_{100}} \approx 3.71$ . Indeed, in Sections 2.5, 4.3.5 and 5.5, our empirical analysis of various series involves analysing the sequence of ratios in order to estimate the growth constant. This is despite not always having a proof that the ratios are increasing, or even converge.

The property of log-convexity often has a natural combinatorial interpretation. For example, if  $\mathbf{a}$  is the counting sequence for a permutation class, then the ratio  $a_n/a_{n-1}$  is the average number of possible ways to insert an element onto the end of a random permutation in the class of size  $n - 1$ , without leaving the permutation class. Hence the counting sequence is log-convex if and only if there are on average more ways to add an element to a permutation of length  $n$  than one of length  $n - 1$ . It is a source of frustration that even in the class of 1324-avoiding permutations this has not been proved, although the weaker statement that the ratios  $a_n/a_{n-1}$  converge in this case was proved by Atapour and Madras [10].

In this chapter we will focus on an even stronger property, that of being a *Stieltjes moment sequence*. We will describe a variety of ways to prove that a sequence is a Stieltjes moment sequence. One result which appears to be new is that if  $\mathbf{a}$  enumerates even-length loops in an undirected, locally finite graph  $\Gamma$ , then it is a Stieltjes moment sequence. As a specific example, this implies that any cogrowth sequence for a Cayley graph is a Stieltjes moment sequence. We also prove a more general result for walks on graphs  $\Gamma$  with weighted edges. It follows from this stronger result that any sequence  $\mathbf{a}$  defined by  $a_n = [t^{2n}]Q(b/t + 1, t^2)$  is also a Stieltjes moment sequence, where  $Q$  is the generating function for weighted quarter-plane walks considered in Chapter 2, and  $b$  is any real number. Finally we will give empirical evidence that some other sequences are Stieltjes moment sequences. In particular we conjecture that the sequence  $(\text{Av}_n(1324))_{n \geq 0}$  is a Stieltjes moment sequence, which would imply a lower bound of 10.302 on its growth rate  $\mu$ , improving on the bound of 10.271 deduced in Chapter 3.

We will give three equivalent definitions of a Stieltjes moment sequence below, but first we define the Hankel matrices of the sequence  $\mathbf{a}$ . For  $n \geq 0$ , the *Hankel matrix*  $H_\infty^{(n)}(\mathbf{a})$  is given by

$$H_\infty^{(n)}(\mathbf{a}) = \begin{bmatrix} a_n & a_{n+1} & a_{n+2} & \cdots \\ a_{n+1} & a_{n+2} & a_{n+3} & \cdots \\ a_{n+2} & a_{n+3} & a_{n+4} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

The following theorem was proven in part by Stieltjes and in part by Gant-

makher and Krein. In particular, the properties (a) and (d) were shown to be equivalent by Stieltjes [196], while these were later shown to be equivalent to (b) and (c) by Gantmakher and Krein [112]. See also [186, 188].

**Theorem 6.1.** *For a sequence  $\mathbf{a} = (a_n)_{n \geq 0}$ , the following are equivalent:*

(a) *There exists a positive measure  $\rho$  on  $[0, \infty)$  such that*

$$a_n = \int x^n d\rho(x).$$

(b) *The matrices  $H_\infty^{(0)}(\mathbf{a})$  and  $H_\infty^{(1)}(\mathbf{a})$  are both positive semidefinite.*

(c) *The matrix  $H_\infty^{(0)}(\mathbf{a})$  is totally positive (all of its minors are non-negative).*

(d) *There exists a sequence of real numbers  $\alpha_0, \alpha_1, \dots \geq 0$  such that the generating function  $A(t)$  for the sequence  $a_0, a_1, \dots$  satisfies*

$$A(t) = \sum_{n=0}^{\infty} a_n t^n = \frac{\alpha_0}{1 - \frac{\alpha_1 t}{1 - \frac{\alpha_2 t}{1 - \dots}}}.$$

**Definition 6.2.** A sequence that satisfies the equivalent conditions of Theorem 6.1 is called a *Stieltjes moment sequence*.

To see that a Stieltjes moment sequence is log-convex, it suffices to observe that for each  $n \geq 1$ , the expression  $a_{n+1}a_{n-1} - a_n^2$  is a minor of  $H_\infty^{(0)}(\mathbf{a})$ , so this expression is non-negative by condition (c). Alternatively, we can deduce log-convexity directly from condition (a) as follows: observe that for any real numbers  $\beta$  and  $\gamma$ , the expression

$$\beta^2 a_{n-1} + 2\beta\gamma a_n + \gamma^2 a_{n+1} = \int x^{n-1} (\beta + \gamma x)^2 d\rho(x)$$

is non-negative. Then setting  $\beta = a_n$  and  $\gamma = -a_{n-1}$  and rearranging yields

$$\frac{a_{n+1}}{a_n} \geq \frac{a_n}{a_{n-1}}.$$

Log convexity of the sequence  $\mathbf{a}$  implies that the ratios  $a_n/a_{n-1}$  are lower bounds for the growth rate  $\mu$ . In the case that  $\mathbf{a}$  is a Stieltjes moment sequence, we will describe a way to compute stronger lower bounds for  $\mu$  in Section 6.1.1.

We now define a more general type of sequence, called a *Hamburger moment sequence*. The following theorem describing equivalent definitions of a Hamburger moment sequence is a classical result regarding the Hamburger moment problem, see for example [1].

**Theorem 6.3.** For a sequence  $\mathbf{a} = (a_n)_{n \geq 0}$ , the following are equivalent:

(a) There exists a positive measure  $\rho$  on  $(-\infty, \infty)$  such that

$$a_n = \int_{-\infty}^{\infty} x^n d\rho(x).$$

(b) The matrix  $H_{\infty}^{(0)}(\mathbf{a})$  is positive semidefinite.

**Definition 6.4.** A sequence that satisfies the equivalent conditions of Theorem 6.3 is called a *Hamburger moment sequence*.

From either definition of a Hamburger moment sequence, it follows immediately that any Stieltjes moment sequence is a Hamburger moment sequence. Some sort of converse also holds: If  $\mathbf{a} = a_0, a_1, \dots$  is a Hamburger moment sequence, then the sub-sequence  $\mathbf{a}^E = a_0, a_2, a_4, \dots$  is a Stieltjes moment sequence. [*Proof.* If  $\mathbf{a}$  is a Hamburger moment sequence then the Hankel matrix  $H_{\infty}^{(0)}(\mathbf{a})$  is positive semidefinite. Deleting all of the even rows and columns in  $H_{\infty}^{(0)}(\mathbf{a})$  yields the matrix  $H_{\infty}^{(0)}(\mathbf{a}^E)$ , while deleting all of the odd rows and columns yields the matrix  $H_{\infty}^{(1)}(\mathbf{a}^E)$ . Hence these are both principal submatrices of  $H_{\infty}^{(0)}(\mathbf{a})$ , so they are also positive semidefinite. Therefore,  $\mathbf{a}^E$  is a Stieltjes moment sequence, using condition (b) in Theorem 6.1.]

Considering the measure  $\rho$  associated to a Hamburger moment sequence  $\mathbf{a}$ , a natural question to ask is: when is this measure uniquely defined by the sequence? A partial answer is given by Carleman's condition [1], which states that the measure  $\rho$  is unique if

$$\sum_{n=0}^{\infty} a_{2n}^{-\frac{1}{2n}} = +\infty.$$

This is certainly true when the sequence grows at most exponentially, as is the case for all our examples. For Stieltjes moment sequences, one only requires the following weaker condition to conclude that the measure  $\rho$  is unique [1]:

$$\sum_{n=0}^{\infty} a_n^{-\frac{1}{2n}} = +\infty.$$

To conclude this introduction, we mention one relationship between the asymptotic behaviour of a Hamburger or Stieltjes moment sequence and its associated measure. For a Hamburger moment sequence  $\mathbf{a}$  given by

$$a_n = \int_{-\infty}^{\infty} x^n d\rho(x)$$

that grows at most exponentially, the radius of convergence of  $A(t) = a_0 + a_1t + a_2t^2 + \dots$  is equal to

$$\frac{1}{|\rho|}.$$

In particular, this means that if  $\mathbf{a}$  is a Stieltjes moment sequence, the exponential growth rate  $\mu$  of  $\mathbf{a}$  is equal to the maximum value in the support of  $\rho$ .

## Bounding the growth rate of Stieltjes moment sequences

One benefit of proving that a sequence  $\mathbf{a}$  is a Stieltjes moment sequence is that it allows us to compute good lower bounds for the exponential growth rate of the sequence using only finitely many terms. The method was described in [128], but we repeat the description here, using the continued fraction form of  $\mathbf{a}$ . Consider the generating function

$$A(t) = \sum_{n=0}^{\infty} a_n t^n = \frac{\alpha_0}{1 - \frac{\alpha_1 t}{1 - \frac{\alpha_2 t}{1 - \dots}}}.$$

Using the terms  $a_0, \dots, a_n$ , we calculate the terms  $\alpha_0, \dots, \alpha_n$ . It is easy to see that the coefficients of  $A(t)$  are nondecreasing in each  $\alpha_j$ . Hence,  $A(t)$  is (coefficientwise) bounded below by the generating function  $A_n(t)$ , defined by setting  $\alpha_{n+1}, \alpha_{n+2}, \dots$  to 0. Therefore, the growth rate  $\mu_n$  of  $A_n(t)$  is no greater than the growth rate  $\mu$  of  $A(t)$ . The growth rates  $\mu_1, \mu_2, \dots$  clearly form a non-decreasing sequence, and, since the coefficients of  $A_n(t)$  are log convex,  $\mu_n \geq a_n/a_{n-1}$ . It follows that this sequence  $\mu_1, \mu_2, \dots$  of lower bounds converges to the exponential growth rate  $\mu$  of  $\mathbf{a}$ .

If we assume further that the sequences  $\alpha_0, \alpha_2, \alpha_4, \dots$  and  $\alpha_1, \alpha_3, \dots$  are non-decreasing, as we find empirically in many of the cases we consider, we can get stronger lower bounds for the growth rate by setting  $\alpha_{n+1}, \alpha_{n+3}, \dots$  to  $\alpha_{n-1}$  and  $\alpha_{n+2}, \alpha_{n+4}, \dots$  to  $\alpha_n$ . For this sequence the exponential growth rate of the corresponding sequence  $\mathbf{a}$  is  $(\sqrt{\alpha_n} + \sqrt{\alpha_{n-1}})^2$ . By the method with which we constructed this bound, it is clear that  $(\sqrt{\alpha_n} + \sqrt{\alpha_{n-1}})^2 \geq \mu_n$ . Hence, the lower bounds  $(\sqrt{\alpha_n} + \sqrt{\alpha_{n-1}})^2$  converge to the growth rate  $\mu$ . In particular, if  $\alpha_n \leq \alpha_{n+2}$  for each  $n$  and the limit

$$\lim_{n \rightarrow \infty} \alpha_n$$

exists, then it is equal to  $\mu/4$ . In fact, in Section 6.3 we find empirically that the sequence  $\alpha_n$  converges to  $\mu/4$  in a number of other sequences which do not satisfy  $\alpha_n \leq \alpha_{n+2}$  for each  $n$ .

# Methods for proving that combinatorial sequences are Stieltjes moment sequences

In this section we describe a variety of methods which can be used to prove that a sequence is a Stieltjes moment sequence. Despite these methods, the number of sequences which we empirically find to be Stieltjes moment sequences goes far beyond what we are able to prove, so there is still much to understand about why so many combinatorial sequences are (at least empirically) Stieltjes moment sequences.

## Loops in graphs and cogrowth sequences

In [128], Haagerup, Haagerup and Ramirez-Solano proved that the cogrowth sequence  $a_0, a_1, \dots$  for Thompson's group  $F$  is the sequence of moments of some probability measure  $\mu$  on  $[0, \infty)$ . In other words, the sequence is a Stieltjes moment sequence. In fact, their proof applies to the cogrowth series of any (locally finite) Cayley graph  $\Gamma$ . In this section, we generalise the result further, to any locally finite graph.

**Theorem 6.5.** *Let  $\Gamma$  be a locally finite graph with a fixed base vertex  $v_0$ . For each  $n \in \mathbb{Z}_{\geq 0}$ , let  $t_n$  be the number of loops of length  $n$  in  $\Gamma$  which start and end at  $v_0$ . Then  $t_0, t_1, \dots$  is a Hamburger moment sequence.*

*Proof.* The sequence  $\mathbf{t} = t_0, t_1, \dots$  is a Hamburger moment sequence if and only if the matrix

$$H_{\infty}^{(0)}(\mathbf{t}) = \begin{bmatrix} t_0 & t_1 & t_2 & \dots \\ t_1 & t_2 & t_3 & \dots \\ t_2 & t_3 & t_4 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

is positive semidefinite. To prove that this matrix is positive semidefinite, we will show that it is the matrix representation of a positive definite bilinear form.

Let  $M$  be the inner product space over  $\mathbb{R}$  defined by the orthonormal basis  $\{b_v | v \in V(\Gamma)\}$ . For each  $n \in \mathbb{Z}$  we let  $x_n \in M$  be the element defined by

$$x_n = \sum_{v \in V(\Gamma)} p_{n,v} b_v,$$

where  $p_{n,v}$  is the number of paths of length  $n$  in  $\Gamma$  from  $v_0$  to  $v$ . Then for any non-negative integers  $m$  and  $n$ , and any vertex  $v$ , the product  $p_{n,v} p_{m,v}$



is equal to the number of  $(m+n)$ -step walks in  $\Gamma$  which start and end at  $v_0$  and whose  $m$ th step ends at  $v$ . Summing over all vertices  $v$  yields the value

$$\langle x_n, x_m \rangle = \sum_{v \in V(\Gamma)} p_{n,v} p_{m,v},$$

which is therefore equal to the number  $t_{m+n}$  of paths of length  $m+n$  in  $\Gamma$  from  $v_0$  to itself. Now let  $X$  be the subspace of  $M$  spanned by  $\{x_0, x_1, \dots\}$ . Then  $H_\infty^{(0)}(\mathbf{t})$  is the matrix representation of the inner product  $\langle, \rangle$ , restricted to  $X$ , with respect to the spanning set  $\{x_0, x_1, \dots\}$ . Therefore,  $H_\infty^{(0)}(\mathbf{t})$  is positive semidefinite. Note that if  $\{x_0, x_1, \dots\}$  are linearly independent, then  $H_\infty^{(0)}(\mathbf{t})$  is positive definite (this occurs if and only if the connected component of  $\Gamma$  containing  $v_0$  is infinite).  $\square$

**Remark.** Since  $\mathbf{t} = t_0, t_1, t_2, \dots$  is a Hamburger moment sequence, it follows that  $\mathbf{t}^E = t_0, t_2, t_4, \dots$  is a Stieltjes moment sequence.

**Remark.** We can generalise this to weighted walks as follows: To each edge  $e$  in  $\Gamma$ , assign a weight  $w_e \in \mathbb{R}$ . Let the weight of a walk be equal to the product of the weights of its steps, where each step is given the weight of the corresponding edge. Then let  $t_n$  be the weighted number of walks of length  $n$  which start and end at  $v_0$ , that is the sum over all such walks  $w$  of the weight of  $w$ . Then the proof still holds, letting  $p_{n,v}$  be the *weighted* number of paths of length  $n$  from  $v_0$  to  $v$ . Hence, the results that  $\mathbf{t} = t_0, t_1, t_2, \dots$  is a Hamburger moment sequence and  $\mathbf{t}^E = t_0, t_2, t_4, \dots$  is a Stieljes moment sequence still hold.

**Remark.** If  $G$  is a group with finite, inverse closed generating set  $S$  and  $\Gamma = \Gamma(G, S)$  is the corresponding Cayley graph, then  $\mathbf{t}^E = t_0, t_2, t_4, \dots$  is the cogrowth sequence for  $(G, S)$ . Hence, the cogrowth sequence of any finitely generated group is a Stieltjes moment sequence. Let  $\rho$  be the corresponding measure on  $(0, \infty)$  which satisfies

$$\int x^n d\rho(x) = t_{2n},$$

for each  $n$ . Then the growth rate  $\mu$  of the sequence  $\mathbf{t}^E$  is equal to the maximum value in the support of  $\rho$ , so  $\rho$  is supported on a subset of  $[0, \mu]$ . Moreover, since  $t_n \leq |S|^n$ , we must have  $\mu \leq |S|^2$ . Hence,  $\rho$  is a measure on  $[0, |S|^2]$ .

**Remark.** Let a *reflected double Kreweras loop* (an *rdk-loop*) be a walk in the quarter-plane starting and ending at the origin with steps from the set  $S = \{(1, 0), (-1, 0), (0, 1), (0, -1), (1, -1), (-1, 1)\}$ . Let  $b$  be a non-negative real number and let  $F(b, t)$  be the weighted generating function for rdk-loops

where  $t$  counts steps and the weight  $b$  is given to diagonal steps.  $F(b, t)$  can be seen to count weighted walks on a graph  $\Gamma$  with vertex set  $\mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$  and an edge between any pair of vertices  $(u, v)$  and  $(u', v')$  satisfying  $(u - u', v - v') \in S$ , where the weight  $b$  assigned to each diagonal edge. Then it follows from Theorem 6.5 that the sequence  $\mathbf{a}$  defined by  $a_n = [t^n]F(b, t)$  is a Hamburger moment sequence. We will now show that  $F(b, t)$  is simply related to  $Q(a, u)$ , where  $Q(a, u)$  is the generating function for weighted quarter-plane walks considered in Chapter 2. Recall that  $Q(a, u)$  enumerates quarter-plane returns to the origin (with unit, axis parallel steps) where  $a$  counts  $NW$  and  $ES$  corners and  $u$  counts the half length. We call the walks counted by  $Q(a, u)$  *axis-parallel loops* (ap-loops). Consider the transformation  $T$  which sends each rdk-loop to an ap-loop by replacing each  $(1, -1)$  step with an  $E$  step followed by an  $S$  step, and each  $(-1, 1)$  step with an  $N$  step followed by a  $W$  step. To reverse this transformation, one needs to consider each  $NW$  and  $ES$  in an ap-loop and either replace it with a diagonal step or no change. Hence, each  $NW$  or  $ES$  corner contributes  $bt + t^2$  to the weights of the corresponding rdk-loops (and  $au$  to the weight of the ap-loop). Each other step in the ap-loop contributes  $\sqrt{u}$  to its weight and  $t$  to the weight of each corresponding rdk-loop. Hence, if  $au = bt + t^2$  and  $\sqrt{u} = t$ , then  $F(b, t) = Q(a, u)$ . In particular,

$$F(b, t) = Q(1 + b/t, t^2).$$

As described in Section 6.1.1, we can compute lower bounds  $\mu_n$  for the exponential growth rate of any such sequence. Turning our attention to Thompson's group, using 31 terms of the cogrowth sequence, we have computed the corresponding terms  $\alpha_0, \alpha_1, \dots, \alpha_{31}$ . Using these we have computed the rigorous lower bound  $\mu_{31} \approx 13.269$  for the exponential growth rate of the cogrowth sequence of Thompson's group. If we assume that the sequences  $\alpha_0, \alpha_2, \dots$  and  $\alpha_1, \alpha_3, \dots$  are increasing, as they appear to be, we get the stronger lower bound  $(\sqrt{\alpha_{30}} + \sqrt{\alpha_{31}})^2 \approx 13.706$ .

## Exactly solved sequences

In this section we consider exactly solved sequences with at most exponential growth. To investigate these sequences we use an alternative characterisation of the Stieltjes property using Pick functions, described by Liu and Pego [150]. A function  $f$  is called a *Pick function* if it is analytic in the upper half plane  $\mathbb{H} = \{z \in \mathbb{C} | \text{Im}(z) > 0\}$  and satisfies  $f(\mathbb{H}) \subseteq \overline{\mathbb{H}}$ .

**Theorem 6.6.** *For a sequence  $\mathbf{a} = a_0, a_1, \dots$  of real numbers and real number  $\tau > 0$ , the following conditions are equivalent:*

(a)  $\mathbf{a}$  is a Stieltjes moment sequence with exponential growth rate at most  $\tau$ .

(b) There exists a positive measure  $\rho$  on  $[0, \tau]$  such that

$$a_n = \int_0^\tau x^n d\rho(x)$$

for each  $n$ .

(c) The analytic continuation of the generating function  $f(z)$  defined by

$$f(z) = \sum_{n=0}^{\infty} a_n z^n,$$

is a Pick function which is analytic and non-negative on  $(-\infty, 1/\tau)$ .

*Proof.* We have already seen that if  $\mathbf{a}$  grows at most exponentially, then the exponential growth rate is equal to the maximum value in the support of  $\rho$ . The equivalence between (a) and (b) follows immediately from this observation. The equivalence of (b) and (c) is due to Corollary 1 in [150].  $\square$

**Remark.** If  $f(z)$  extends to a continuous function on  $\overline{\mathbb{H}} \cup \{\infty\}$  which is analytic in  $\mathbb{H}$ , one can verify that  $f(z)$  is a Pick function by checking that  $f(\mathbb{R}) \subset \overline{\mathbb{H}}$  [*Proof.*  $\overline{\mathbb{H}} \cup \{\infty\}$  is a compact set, so its image must also be compact. Hence  $f(\mathbb{H})$  is bounded. Moreover, by the open mapping theorem, the image  $f(\mathbb{H})$  must be open. Hence, the boundary of  $f(\overline{\mathbb{H}} \cup \{\infty\})$  is contained in  $f(\mathbb{R} \cup \{\infty\}) = \overline{f(\mathbb{R})}$ , which is a subset of  $\overline{\mathbb{H}}$ . Hence  $f(\mathbb{H}) \subseteq \mathbb{H}$ .]

**Remark.** If the measure  $\rho$  is differentiable, then it has a density function  $u(x) = \rho'(x)$ . We will describe how to calculate  $u(x)$  in terms of the function  $f$  in these cases. The function  $f(z)$  satisfies

$$f(z) = \int_0^\tau \frac{1}{1-xz} d\rho(x) = \int_0^\tau \frac{u(x)}{1-xz} dx.$$

Setting  $z = 1/y + ci$  for positive real numbers  $c$  and  $y$ , then taking the imaginary part yields

$$\operatorname{Im}(f(1/y + ci)) = \int_0^\tau \frac{cxu(x)}{(1-x/y)^2 + c^2x^2} dx.$$

Taking the limit as  $c \rightarrow 0$  then dividing by  $\pi y$  yields

$$u(y) = \frac{1}{\pi y} \operatorname{Im}(f(1/y)),$$

as long as  $u(x)$  is continuous at  $x = y$ .

Now we will give a number of examples

**Example 6.7.** The Catalan numbers  $C_n = \frac{1}{n+1} \binom{2n}{n}$  have the counting function

$$f(z) = \sum_{n=0}^{\infty} C_n z^n = \frac{1 - \sqrt{1 - 4z}}{2z}.$$

This function is analytic on  $\mathbb{H}$  and continuous on  $\overline{\mathbb{H}}$ . Moreover,  $f(z) \rightarrow 0$  as  $z \rightarrow \infty$ , so  $f(z)$  is continuous on  $\overline{\mathbb{H}} \cup \{\infty\}$ . Hence, using our first remark on Theorem 6.6, along with the fact that  $f$  sends  $\mathbb{R}$  to a subset of  $\overline{\mathbb{H}}$ , the function  $f(z)$  is a Pick function. We can also observe that  $f(z)$  is analytic and non-negative on  $(-\infty, 1/4)$ , so the Catalan numbers form a Stieltjes moment sequence. Using the second remark, for  $x \in (0, 4)$ , the density function  $u$  is given by

$$u(x) = \frac{1}{\pi x} \operatorname{Im}(f(1/x)) = \frac{\sqrt{4-x}}{2\pi\sqrt{x}}.$$

Note that we can easily verify that this is the density function as

$$C_n = \frac{1}{n+1} \binom{2n}{n} = \int_0^4 x^n \frac{\sqrt{4-x}}{2\pi\sqrt{x}} dx = \int_0^\infty x^n u(x) dx.$$

**Example 6.8.** Let  $a_n = Av_n(1342)$  be the number of 1342 avoiding permutations of length  $n$ . We know that the counting function

$$f(z) = \sum_{n=0}^{\infty} a_n z^n = \frac{32z}{1 + 20z - 8z^2 - (1 - 8z)^{3/2}}.$$

As in the previous example, we can check that  $f(z)$  is a Pick function which is analytic on  $(-\infty, 1/8)$ . Using the second remark we can calculate the density function

$$u(x) = \frac{1}{\pi x} \operatorname{Im}(f(1/x)) = \frac{(8-x)^{3/2}\sqrt{x}}{2\pi(x+1)^3},$$

for  $x \in (0, 8)$ . Again, we can easily verify that this density function corresponds to the sequence given by  $a_n = Av_n(1342)$ .

In each of these examples we had an exact *algebraic* expression for the counting function. This method can also apply in other cases in which the counting function  $f(x)$  is exactly solved, however in general it is not trivial to prove that  $f(x)$  is a Pick function. For example, a number of years ago, Alan Sokal conjectured that the Apéry numbers

$$A_n = \sum_{k=0}^n \binom{n+k}{k}^2 \binom{n}{k}^2$$

form a Stieltjes moment sequence [82]. This was only very recently proved by Gerald Edgar, as he gave an explicit formula for the corresponding density function in terms of Heun functions [83]. Sokal has Conjectured more generally that for any constant  $c \geq 1$ , the sequence defined by

$$A_n(c) = \sum_{k=0}^n \binom{n+k}{k}^2 \binom{n}{k}^2 c^k,$$

is a Stieltjes moment sequence [188], but this has not been proved for any  $c > 1$ .

## Continued fraction approach

For a number of sequences  $\mathbf{a} = a_0, a_1, \dots$  it is possible to find an explicit form of the coefficients  $\alpha_0, \alpha_1, \alpha_2, \dots$  in the continued fraction expansion

$$A(t) = \sum_{n=0}^{\infty} a_n t^n = \frac{\alpha_0}{1 - \frac{\alpha_1 t}{1 - \frac{\alpha_2 t}{1 - \dots}}}.$$

We call the sequence  $\alpha_0, \alpha_1, \dots$  the *alpha sequence* of  $\mathbf{a}$ . Then it suffices to check that each  $\alpha_i$  is positive in order to deduce that the sequence  $\mathbf{a} = a_0, a_1, \dots$  is a Stieltjes moment sequence. For example, if  $\mathbf{a}$  is the sequence of Catalan numbers, then  $\alpha_k = 1$  for all  $k$ . For the factorial sequence  $a_n = n!$ , the coefficients  $\alpha_k$  are given by  $\alpha_0 = 1$  and  $\alpha_{2k-1} = \alpha_{2k} = k$  for all positive integers  $k$ .

We will see a variety of other examples in Chapter 7. We will now show that for  $a_0 = 1$ , this is in fact a specialisation of Theorem 6.5. Given a sequence  $\mathbf{a}$  and the corresponding alpha sequence  $\alpha_0, \alpha_1, \dots$ , which satisfies  $\alpha_0 = 1$  and  $\alpha_n > 0$  for each  $n$ , we construct a weighted graph  $\Gamma$  as follows: The vertex set of  $\Gamma$  is denoted  $\{v_k | k \in \mathbb{Z}_{\geq 0}\}$ , and for each  $k \in \mathbb{Z}_{>0}$ , there is an edge  $e_k$  between  $v_{k-1}$  and  $v_k$  with weight  $\sqrt{\alpha_k}$ . Now let  $B_k(t)$  denote the generating function for weighted walks in  $\Gamma$  starting and ending at  $v_k$  and never visiting  $v_{k-1}$ , counted by half-length. Let  $A_k(t)$  denote the generating function for the subset of these walks with length at least 1, which only visit  $v_k$  at the start and end. Then the walks counted by  $B_k$  are the same as those counted by  $A_{k+1}$ , with an extra step along edge  $e_{k+1}$  at the start and end. So  $B_k(t) = \alpha_{k+1} t A_{k+1}(t)$ . Moreover, each path counted by  $A_k(t)$  is a unique sequence of paths counted by  $B_k(t)$ , so

$$A_k(t) = \frac{1}{1 - B_k(t)} = \frac{1}{1 - \alpha_{k+1} t A_{k+1}(t)}.$$

Iterating this equality shows that

$$A_0(t) = \frac{\alpha_0}{1 - \frac{\alpha_1 t}{1 - \frac{\alpha_2 t}{1 - \dots}}},$$

as  $\alpha_0 = 1$ .

Recently, Pétréolle, Sokal and Zhu have found a generalised version of continued fractions, called *branched* continued fractions, which they show also yield Stieltjes moment sequences, as long as their coefficients are positive [166]. In terms of graphs, they prove the following:

**Theorem 6.9.** *Let  $m$  be a fixed positive integer and let  $\Gamma$  be a weighted directed graph, with vertex set  $\{v_k | k \in \mathbb{Z}_{\geq 0}\}$ , and edges with non-negative weights from  $v_k$  to  $v_{k+1}$  and  $v_{k+m}$  to  $v_k$  for each  $m$ . For each  $n \in \mathbb{Z}_{\geq 0}$ , let  $a_n$  be the weighted number of walks of length  $(m+1)n$  in  $\Gamma$ , starting and ending at  $v_0$ . Then the sequence  $a_0, a_1, a_2, \dots$  is a Stieltjes moment sequence.*

By choosing specific values of the weights of the edges, Pétréolle, Sokal and Zhu prove that a variety of sequences are Stieltjes moment sequences. For example, they show that the reversed second-order Eulerian polynomials, defined by  $P_0(x) = 1$  and

$$P_n(x) = (x - x^2)P'_{n-1}(x) + (n + (n-1)x)P_{n-1}(x)$$

form a Stieltjes moment sequence for any  $x \geq 0$ . In fact they prove the stronger result that the sequences of polynomials  $P_0(x), P_1(x), \dots$  is Hankel totally positive. We discuss Hankel total positivity further in Section 6.4, and we discuss the reversed second order Eulerian polynomials further in Chapter 7, where we denote them by  $\overline{E}_n^{[2]}(x)$ .

Since branched continued fractions have a natural interpretation as weighted walks on a graph, one may hope to unify this with Theorem 6.5 by finding a general class of weighted directed graphs on which the enumeration of weighted walks starting and ending at a fixed base vertex  $v_0$  form a Stieltjes moment sequence.

## Permutations avoiding increasing patterns

The numbers  $\text{Av}_n(12\dots k)$  have already been described as moments of a probability distribution by Rains [173]. In particular he showed that the number  $a_n = \text{Av}_n(12\dots k)$  of permutations of length  $n$  which contain no increasing subsequence of length  $k$  is equal to the  $n$ th moment of the modulus

squared of the trace of a  $(k-1) \times (k-1)$  Haar random unitary matrix. Since  $a_n$  is the  $n$ th moment of some random variable  $X$  taking values in  $\mathbb{R}_{\geq 0}$ , the sequence  $a_0, a_1, \dots$  is a Stieltjes moment sequence.

## Empirical Stieltjes moment sequences

Although there are relatively few sequences that we can prove to be Stieltjes moment sequences, we have found empirically that a number of other combinatorial sequences discussed in this thesis appear to be Stieltjes moment sequences. Our method for testing this is quite simple - given the terms  $a_0, a_1, \dots, a_n$  of the sequence, we can calculate the corresponding coefficients  $\alpha_0, \alpha_1, \dots, \alpha_n$ . If any of these coefficients  $\alpha_i$  is negative, then the sequence  $\mathbf{a}$  is not a Stieltjes moment sequence. On the other hand, if all of these coefficients are positive, it suggests that the sequence may be a Stieltjes moment sequence, especially if we have a large number of terms. In this case we can refine our analysis by analysing the sequence  $\alpha_0, \alpha_1, \dots, \alpha_n$ , and trying to predict its asymptotic behaviour. In some cases, most notably the sequence  $\text{Av}_n(1324)$ , we find compelling evidence that this sequence is converging, which leads us to conjecture that these are Stieltjes moment sequences. In order to observe some possible behaviour of Stieltjes moment sequences, we start by analysing the alpha sequence of some combinatorial sequences which we know are Stieltjes moment sequences.

## Empirical analysis of known Stieltjes moment sequences

In this section we analyse the alpha sequences  $\alpha_0, \alpha_1, \dots$  of a number of Stieltjes moment sequences considered in this thesis. We will use the knowledge gained from this to help us analyse sequences which are not known to be Stieltjes moment sequences.

We begin with the two sequences  $(\text{Av}_n(1342))_{n \geq 0}$  and  $(\text{Av}_n(1234))_{n \geq 0}$ . Recall that  $(\text{Av}_n(1342))_{n \geq 0}$  is a Stieltjes moment sequence as shown in Example 6.8, while  $(\text{Av}_n(1234))_{n \geq 0}$  is a Stieltjes moment sequence as discussed in Section 6.2.4. The corresponding alpha sequences are plotted against  $1/n$  in Figures 6.1 and 6.2. Recall that the growth rates  $\mu_{1342}$  and  $\mu_{1234}$  of these sequences are 8 and 9 respectively. In each case, it appears that the alpha sequence converges to  $\mu/4$ , with the even coefficients  $\alpha_{2k}$  converging from below and the odd coefficients  $\alpha_{2k+1}$  converging from above.

Now we consider the cogrowth sequences of five groups: the Heisenberg group  $H$ , the wreath product  $\mathbb{Z} \wr \mathbb{Z}$ , the iterated wreath product  $\mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z}$ , the Brin-Navas group  $B$  and Thompson's group  $F$ . We know from Theorem 6.5

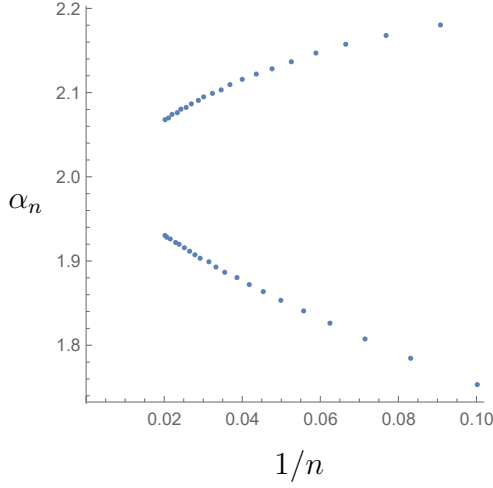


Figure 6.1: Plot of  $\alpha_n$  vs.  $1/n$  for the sequence  $\text{Av}_n(1342)$ , using  $n \in [10, 50]$ .

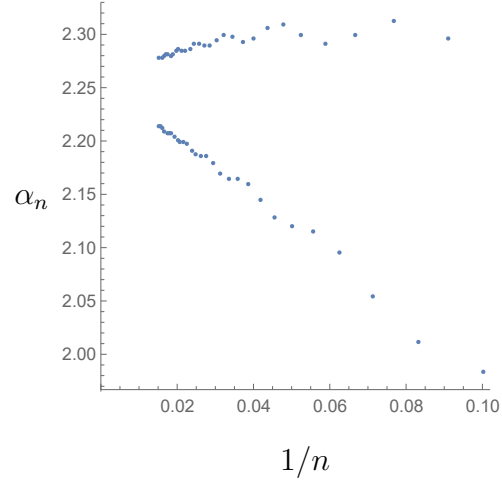


Figure 6.2: Plot of  $\alpha_n$  vs.  $1/n$  for the sequence  $\text{Av}_n(1234)$ , using  $n \in [10, 50]$ .

that these sequences are all Stieltjes moment sequences, so their corresponding alpha sequences have non-negative terms. A plot of  $\alpha_n$  against  $1/n$  for the Heisenberg group is shown in Figure 6.3. Based on this figure it seems likely that the alpha sequence converges to  $4 = \mu/4$ , but this is by no means a certainty. It is perhaps surprising that this alpha sequence is so poorly behaved considering that the corresponding generating function

$$A(t) = \sum_{k=0}^{\infty} a_k t^k = \frac{\alpha_0}{1 - \frac{\alpha_1 t}{1 - \frac{\alpha_2 t}{1 - \dots}}}$$

has a simple power law singularity.

A plot of  $\alpha_n$  against  $\log(n)/n$  for  $\mathbb{Z} \wr \mathbb{Z}$  is shown in Figure 6.4. This plot appears to be roughly linear, suggesting that  $\alpha_n - c_0 \sim c_1 \log(n)^\gamma n^\delta$  for  $\gamma \approx 1, \delta \approx -1$  and some constants  $c_0$  and  $c_1$ . Due to the deviation of the terms it is difficult to ascertain the exact values of  $\gamma$  and  $\delta$ , if they exist. In any case, it appears that the sequence  $\alpha_n$  is converging from below to some constant  $c_0 \approx 4$ . Since the growth rate  $\mu$  of the cogrowth sequence of  $\mathbb{Z} \wr \mathbb{Z}$  is 16, if the alpha sequence is converging from below, then the limit point must be exactly  $\mu/4 = 4$ . For the cogrowth sequence of  $\mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z}$ , the continued fraction coefficients  $\alpha_n$  are plotted against  $\log(n)^{2/3} n^{-2/3}$  in Figure 6.5. This plot appears to be linear, and a limit of 9 seems plausible, suggesting that  $\alpha_n - 9 \sim c \log(n)^{2/3} n^{-2/3}$  for some constant  $c \approx -5$ . Moreover, it appears



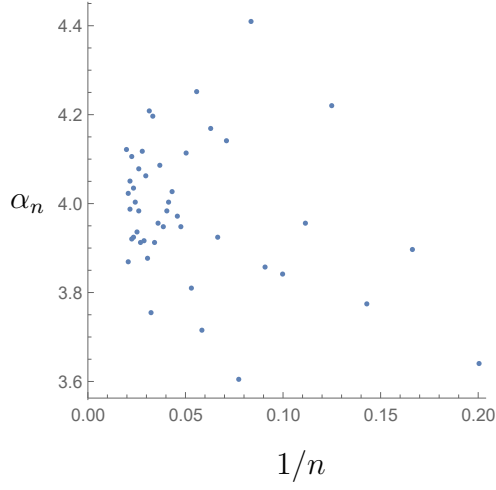


Figure 6.3: Plot of  $\alpha_n$  vs.  $1/n$  for the cogrowth sequence of Heisenberg's group, using  $n \in [5, 50]$ .

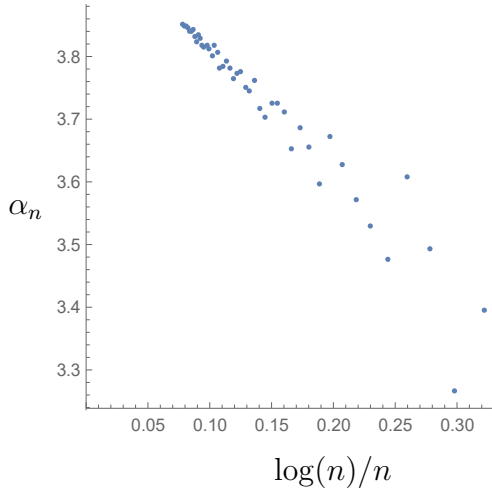


Figure 6.4: Plot of  $\alpha_n$  vs.  $\log(n)/n$  for the cogrowth sequence of  $\mathbb{Z} \wr \mathbb{Z}$ , using  $n \in [5, 50]$ .

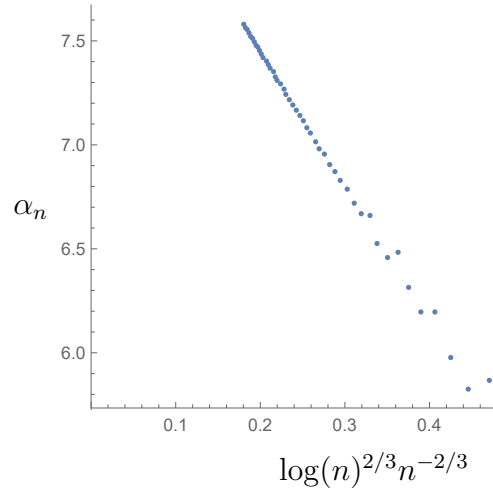


Figure 6.5: Plot of  $\alpha_n$  vs.  $\log(n)^{2/3} n^{-2/3}$  for the cogrowth sequence of  $\mathbb{Z} \wr \mathbb{Z} \wr \mathbb{Z}$ , using  $n \in [5, 50]$ .

that the sequence  $\alpha_n$  is increasing for  $n \geq 12$ , which would imply that  $\alpha_n \rightarrow 9$  as  $n \rightarrow \infty$ .

In Figure 6.6, the values  $\alpha_n$  for the Brin-Navas group  $B$  are plotted against  $1/n$ . We have not been able to guess the asymptotic behaviour of this sequence, which is perhaps unsurprising, since we could not guess the asymp-

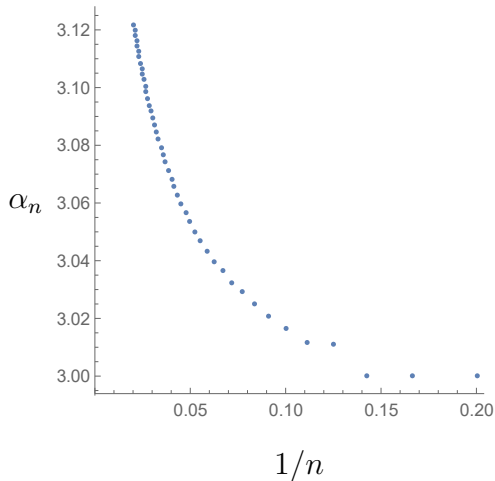


Figure 6.6: Plot of  $\alpha_n$  vs.  $1/n$  for the cogrowth sequence of the Brin-Navas group  $B$ , using  $n \in [5, 50]$ .

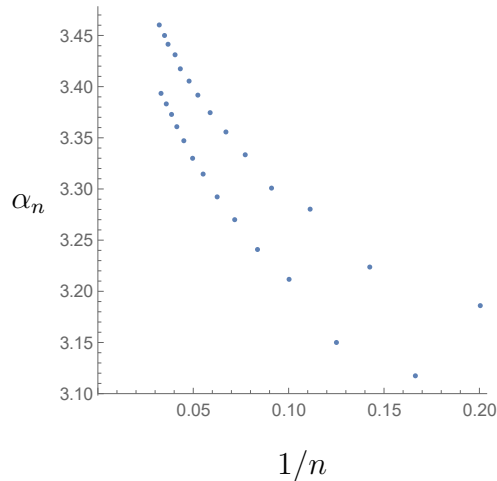


Figure 6.7: Plot of  $\alpha_n$  vs.  $1/n$  for the cogrowth sequence of Thompson's group  $F$ , using  $n \in [5, 31]$ .

otic form of the cogrowth sequence for the Brin-Navas group  $B$ . Assuming that the sequence  $\alpha_n$  is non-decreasing, it must converge to 4, since the cogrowth sequence for  $B$  has growth rate 16.

In Figure 6.6, the values  $\alpha_n$  for Thompson's group  $F$  are plotted against  $1/n$ . It appears that the even terms  $\alpha_{2n}$  and the odd terms  $\alpha_{2n+1}$  each form increasing subsequences, in other words  $\alpha_{n+2} \geq \alpha_n$  for each  $n$ . Assuming this holds for all  $n$ , each of the two subsequences must converge, say  $\alpha_{2m} \rightarrow c_0$  and  $\alpha_{2n+1} \rightarrow c_1$ . Moreover, the growth rate  $\mu$  of the cogrowth sequence for Thompson's group is equal to  $(\sqrt{c_0} + \sqrt{c_1})^2$ . It seems plausible that  $c_0 = c_1 \approx 3.75$ , corresponding to the growth rate  $\mu \approx 15$  predicted in Section 5.5.

## Pattern avoiding permutations

We have already seen that the sequence  $\text{Av}_n(1342)$  is a Stieltjes moment sequence, as is the sequence  $\text{Av}_n(12 \dots k)$  for each  $k$ . Due to the Wilf equivalences of short patterns, this implies that for any permutation pattern  $\pi$  of length 2, 3 or 4, apart from 1324 and 4231, the sequence  $\text{Av}_n(\pi)$  is a Stieltjes moment sequence. A natural question to ask then is the following

**Question.** Is it the case that for any permutation  $\pi$  of length at least 2, the sequence  $\text{Av}_n(\pi)$  is a Stieltjes moment sequence?

Unfortunately for other patterns  $\pi$  of length greater than 4 the existing computation of the numbers  $\text{Av}_n(\pi)$  is very limited. For each pattern of

length 5, apart from patterns which are Wilf-equivalent to the increasing pattern, only the first 13 terms  $\text{Av}_n(\pi)$  are known (see sequences A116485 and A256195-A256208 in [198]). While we have found no counter examples to the above question, this may just be a case of insufficient terms, so it would seem optimistic to conjecture an affirmative answer to this question even in the case of patterns of length 5.

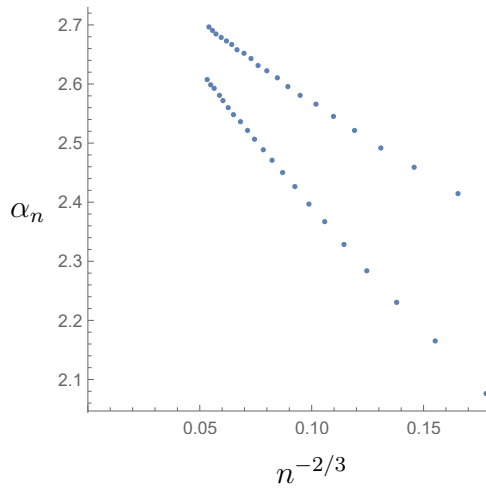


Figure 6.8: Plot of  $\alpha_n$  vs.  $n^{-2/3}$  for the sequence  $\text{Av}_n(1324)$ , using  $n \in [10, 50]$ .

We now turn our attention to the sequence  $\text{Av}_n(1324)$ , which Conway, Guttmann and Zinn-Justin have computed up to  $n = 50$ . The corresponding continued fraction coefficients  $\alpha_n$  are plotted against  $n^{-2/3}$  in Figure 6.8. It appears that the even coefficients  $\alpha_{2k}$  and the odd coefficients  $\alpha_{2k+1}$  each form an increasing sequence which is roughly linear when plotted against  $n^{-2/3}$ . If this trend continues, the terms  $\alpha_i$  will certainly all be positive. Based on this evidence we make the following conjecture:

**Conjecture 6.10.** *The sequence  $(\text{Av}_n(1324))_{n \geq 0}$  is a Stieltjes moment sequence.*

If this could be proven, it would imply a lower bound of 10.302 on the growth rate  $\mu$  of 1324-avoiding permutations, using the method described in Section 6.1.1 applied to the terms  $\text{Av}_n(1324)$  for  $n \leq 50$ , computed by Conway, Guttmann and Zinn-Justin [71]. If we assume further that the sequences  $(\alpha_{2n})_{n \geq 0}$  and  $(\alpha_{2n+1})_{n \geq 0}$  are both increasing, we get an improved lower bound on  $\mu$  of 10.607. Either of these would improve on the best proven lower bound 10.271, which we deduced in Chapter 3.

Extrapolating the two subsequences  $(\alpha_{2n})_{n \geq 0}$  and  $(\alpha_{2n+1})_{n \geq 0}$ , it seems plausible that these both converge to the same constant  $c \approx 2.9$ , which is consistent with the growth rate  $\mu \approx 11.6$  predicted by Conway, Guttmann and Zinn-Justin [71].

## Eulerian orientations

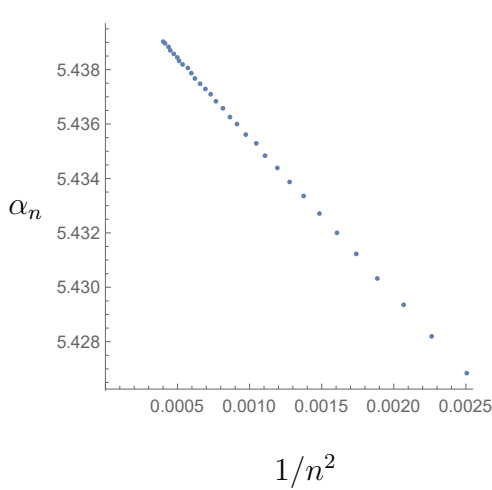


Figure 6.9: Plot of  $\alpha_n$  vs.  $1/n^2$  for the sequence  $q_n$  enumerating quartic Eulerian orientations, using  $n \in [20, 50]$ .

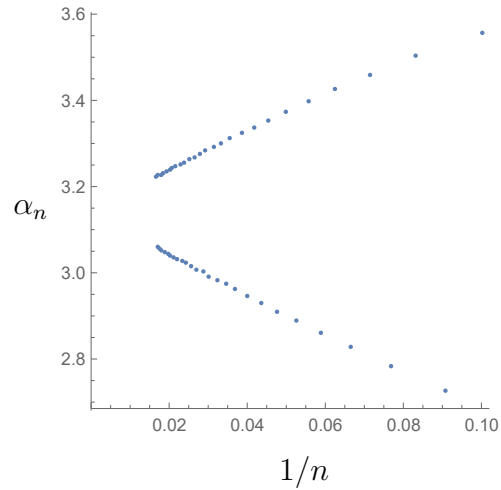


Figure 6.10: Plot of  $\alpha_n$  vs.  $1/n$  for the sequence  $g_n$  enumerating Eulerian orientations, using  $n \in [10, 50]$ .

We now analyse the two sequences considered in Chapter 4, namely  $\mathbf{q} = q_0, q_1, q_2, \dots$  and  $\mathbf{g} = g_0, g_1, g_2, \dots$ , where  $q_n$  is the number of quartic planar Eulerian orientations with  $n$  vertices and  $g_n$  is the number of planar Eulerian orientations with  $n$  edges. The continued fraction coefficients  $\alpha_n$  for quartic planar Eulerian orientations are plotted against  $1/n^2$  in Figure 6.9, while the coefficients for general Eulerian orientations are plotted against  $1/n$  in Figure 6.10. In both of these plots it clearly appears that the terms  $\alpha_n$  are all positive, so we conjecture that both sequences  $\mathbf{q}$  and  $\mathbf{g}$  are Stieltjes moment sequences. Since we know from Theorems 4.1 and 4.2 that the corresponding generating functions  $Q(t)$  and  $G(t)$  are D-algebraic, it may be possible to prove that  $\mathbf{q}$  and  $\mathbf{g}$  are Stieltjes moment sequences using Theorem 6.6, though so far we have not been able to do this.

For the alpha sequence corresponding to  $\mathbf{g}$ , the limit  $\alpha_n \rightarrow \pi$  seems plausible, which is expected since the growth rate in this case is  $\mu = 4\pi$ . For the alpha sequence corresponding to  $\mathbf{q}$ , the plot of  $\alpha_n$  against  $1/n^2$  is visually

linear, so we can estimate the limit quite accurately by taking the intersection of this line with the vertical axis. This yields a limit of approximately 5.441, which corresponds to  $\mu \approx 21.764$ . To get a more accurate estimate of  $\mu/4$ , we can take the linear intercept in this graph formed by successive points  $(1/n^2, \alpha_n)$  and  $(1/(n+1)^2, \alpha_{n+1})$ . The corresponding  $\mu$  estimates are plotted against  $1/n^2$  in Figure 6.11, and they convincingly pass through the origin, which is exactly  $4\sqrt{3}\pi = 21.76559\dots$ . This serves as a demonstration of the general possibility that using the alpha sequence can provide a way to accurately estimate the growth constant of sequences which are at least believed to be Stieltjes moment sequences.

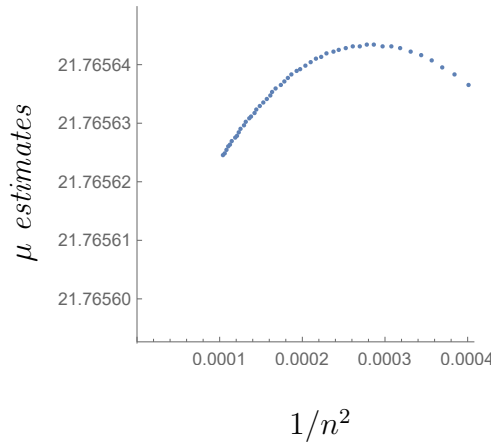


Figure 6.11: Plot of estimates of  $\mu$  vs.  $1/n^2$  for the sequence  $\mathbf{g}$ , using  $n \in [50, 100]$ .

## Machine sortable permutations

We now analyse the three sequences considered in Chapter 2, which enumerate 2sip-sortable permutations, deque-sortable permutations and 2sis-sortable permutations. The terms  $\alpha_n$  for  $n \in [5, 50]$  are plotted against  $1/n$  for 2sip-sortable permutations and deque-sortable permutations in Figures 6.12 and 6.13, respectively. although these coefficients  $\alpha_n$  are all positive, this serves as a warning, as for 2sip-sortable permutations, the coefficient  $\alpha_{55}$  is negative, while for deque-sortable permutations, the coefficient  $\alpha_{66}$  is negative. For 2sis-sortable permutations, the coefficients  $\alpha_n$  are positive for  $n \leq 19$ , but their behaviour, shown in Figure 6.14, does not clearly suggest that they will remain positive. Hence, it is possible that the counting sequence of 2sis-sortable permutations is a Stieltjes moment sequence,

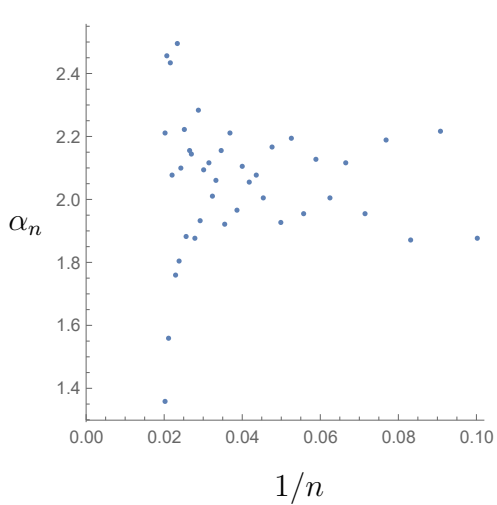


Figure 6.12: Plot of  $\alpha_n$  vs.  $1/n$  for the sequence  $p_n$  enumerating 2sip-sortable permutations, using  $n \in [10, 50]$ .

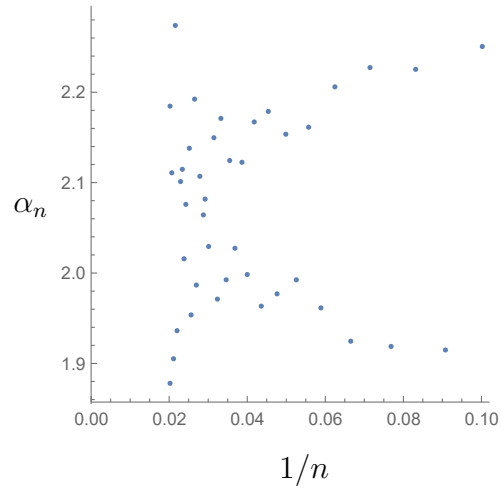


Figure 6.13: Plot of  $\alpha_n$  vs.  $1/n$  for the sequence  $d_n$  enumerating deque-sortable permutations, using  $n \in [10, 50]$ .

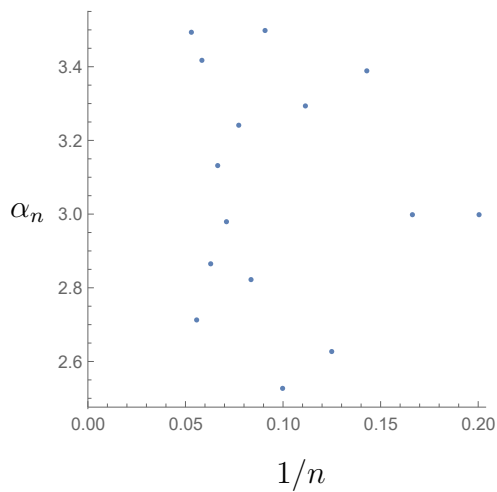


Figure 6.14: Plot of  $\alpha_n$  vs.  $1/n$  for the sequence  $s_n$  enumerating 2sis-sortable permutations, using  $n \in [5, 19]$ .

however it is perhaps more likely that the alpha sequence for 2sis-sortable permutations behaves similarly to that of 2sip-sortable permutations and deque-sortable permutations, in which case this counting sequence is not a Stieltjes moment sequence.

## Hankel total positivity

In this section we briefly mention a generalisation of Stieltjes moment sequences to sequences of polynomials in one or more variables.

**Definition 6.11.** A sequence  $\mathbf{P}$  of polynomials  $P_0, P_1, \dots$  is called *Hankel totally positive* if all minors of the Hankel matrix

$$H_{\infty}^{(0)}(\mathbf{P}) = \begin{bmatrix} P_0 & P_1 & P_2 & \dots \\ P_1 & P_2 & P_3 & \dots \\ P_2 & P_3 & P_4 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

are coefficientwise positive.

If  $\mathbf{P}$  is a sequence of real numbers, then this definition is equivalent to being a Stieltjes moment sequence, as it implies that the Hankel matrix is totally positive in the usual sense. Recently a number of authors have found sequences of polynomials which are Hankel totally positive [166, 189] using continued fraction techniques. Sokal has found many other sequences of polynomials which, empirically, seem to be Hankel totally positive [186, 188]. In Chapter 7 we will describe a sequence of polynomials for perfect matchings in infinitely many variables which we show to be Hankel totally positive.

# Chapter 7

## Phylogenetic trees, augmented perfect matchings and a Thron-type continued fraction (T-fraction) for the Ward polynomials

### Introduction

If  $(a_n)_{n \geq 0}$  is a sequence of combinatorial numbers or polynomials with  $a_0 = 1$ , it is often fruitful to seek to express its ordinary generating function as a continued fraction. The most commonly studied types of continued fractions are Stieltjes-type (S-fractions),

$$\sum_{n=0}^{\infty} a_n t^n = \frac{1}{1 - \frac{\alpha_1 t}{1 - \frac{\alpha_2 t}{1 - \dots}}} \quad , \quad (7.1)$$

and Jacobi-type (J-fractions),

$$\sum_{n=0}^{\infty} a_n t^n = \frac{1}{1 - \gamma_0 t - \frac{\beta_1 t^2}{1 - \gamma_1 t - \frac{\beta_2 t^2}{1 - \dots}}} \quad . \quad (7.2)$$

(Both sides of these expressions are to be interpreted as formal power series in the indeterminate  $t$ .) This line of investigation goes back at least to Euler



[98, 99], but it gained impetus following Flajolet's [104] seminal discovery that any S-fraction (resp. J-fraction) can be interpreted combinatorially as a generating function for Dyck (resp. Motzkin) paths with suitable weights for each rise and fall (resp. each rise, fall and level step). There are now dozens of sequences  $(a_n)_{n \geq 0}$  of combinatorial numbers or polynomials for which a continued-fraction expansion of the type (7.1) or (7.2) is explicitly known.

A less commonly studied type of continued fraction is the Thron-type (T-fraction):

$$\sum_{n=0}^{\infty} a_n t^n = \frac{1}{1 - \delta_1 t - \frac{\alpha_1 t}{1 - \delta_2 t - \frac{\alpha_2 t}{1 - \delta_3 t - \frac{\alpha_3 t}{1 - \dots}}}}. \quad (7.3)$$

Clearly the T-fractions are a generalization of the S-fractions, and reduce to them when  $\delta_i = 0$  for all  $i$ . Besides the S-fractions, there are four special types of T-fractions that are comparatively simple:

- 1) Euler [97, Chapter 18] (see also [208, pp. 17–18]) showed that an arbitrary sequence of non-zero real numbers (or finite sequence of non-zero real numbers followed by zeros) can be written as a special type of T-fraction, namely, one in which  $\delta_1 = 0$  and  $\delta_i = -\alpha_{i-1}$  for  $i \geq 2$ :

$$\frac{1}{1 - \frac{\alpha_1 t}{1 + \alpha_1 t - \frac{\alpha_2 t}{1 + \alpha_2 t - \frac{\alpha_3 t}{1 - \dots}}}} = \sum_{n=0}^{\infty} \alpha_1 \alpha_2 \cdots \alpha_n t^n. \quad (7.4)$$

- 2) If  $\delta_n = 0$  for all *even*  $n$ , then the T-fraction reduces to a J-fraction via the contraction formula [188]

$$\gamma_0 = \alpha_1 + \delta_1, \quad (7.5a)$$

$$\gamma_n = \alpha_{2n} + \alpha_{2n+1} + \delta_{2n+1}, \quad \text{for } n \geq 1, \quad (7.5b)$$

$$\beta_n = \alpha_{2n-1} \alpha_{2n}. \quad (7.5c)$$

This generalises a well-known contraction formula for S-fractions [208, p. 21] [207, p. V-31].

- 3) If  $\delta$  is periodic of period 2 — that is,  $\delta_{2k-1} = x$  and  $\delta_{2k} = y$  — then the sequence  $(a_n)_{n \geq 0}$  generated by the T-fraction is a linear transform of the sequence generated by the S-fraction with the same coefficients  $\alpha$ : see [16, Propositions 3 and 15] for some special cases, and [188] for the general case.
- 4) If  $\delta_n = 0$  for all  $n \geq k$  for some fixed  $k$ , then the sequence  $(a_n)_{n \geq 0}$  generated by the T-fraction is a rational transform of the sequence generated by the S-fraction with the same coefficients  $\alpha$  [188].

But T-fractions not of these four special types are comparatively rare in the combinatorial literature. It is our purpose here to provide a non-trivial example.

Two decades ago, Roblet and Viennot [178] showed that the general T-fraction (7.3) can be interpreted combinatorially as a generating function for Dyck paths in which falls from peaks and non-peaks get different weights. When  $\delta = \mathbf{0}$  these two weights coincide, and the Roblet–Viennot formula reduces to Flajolet’s interpretation of S-fractions in terms of Dyck paths. More recently, several authors [110, 164, 133, 188] have independently found an alternate (and simpler) combinatorial interpretation of the general T-fraction (7.3): namely, as a generating function for Schröder paths with suitable weights for each rise, fall and long level step. When  $\delta = \mathbf{0}$  the long level steps get zero weight, and this formula again reduces to Flajolet’s interpretation of S-fractions in terms of Dyck paths. We will review this Schröder-path representation in Section 7.3.1.

Our combinatorial example concerns the Ward polynomials  $W_n(x)$ , which are defined as follows. Firstly, the *Ward numbers*  $W(n, k)$  [209] are defined by the linear recurrence

$$W(n, k) = k W(n - 1, k) + (n + k - 1) W(n - 1, k - 1) \quad (7.6)$$

with initial condition  $W(0, k) = \delta_{0k}$ . The triangular array of Ward numbers

begins as										Row sums
$n \setminus k$	0	1	2	3	4	5	6	7	8	
0	1									1
1	0	1								1
2	0	1	3							4
3	0	1	10	15						26
4	0	1	25	105	105					236
5	0	1	56	490	1260	945				2752
6	0	1	119	1918	9450	17325	10395			39208
7	0	1	246	6825	56980	190575	270270	135135		660032
8	0	1	501	22935	302995	1636635	4099095	4729725	2027025	12818912

The row sums are [198, A000311]. It is easy to see that  $W(n, n) = (2n - 1)!!$  and that for  $n \geq 1$  we have  $W(n, 0) = 0$ ,  $W(n, 1) = 1$  and  $W(n, 2) = 2^{n+1} - n - 3$ . We then define the *Ward polynomials* to be the row-generating polynomials of this triangular array:

$$W_n(x) = \sum_{k=0}^n W(n, k) x^k. \quad (7.7)$$

For some purposes it is convenient to use instead the reversed Ward polynomials  $\bar{W}_n(x) = x^n W_n(1/x)$ , which satisfy  $\bar{W}_n(0) = (2n - 1)!!$ . See [209, 66, 14, 15] for further information on the Ward numbers and Ward polynomials.

Our first result is:

**Theorem 7.1.** *The ordinary generating function of the Ward polynomials can be expressed by the T-fraction*

$$\sum_{n=0}^{\infty} W_n(x) t^n = \frac{1}{1 - \frac{xt}{1 - t - \frac{2xt}{1 - 2t - \frac{3xt}{1 - \dots}}}} \quad (7.8)$$

with coefficients  $\delta_i = i - 1$  and  $\alpha_i = ix$ . Equivalently, the ordinary generating function of the reversed Ward polynomials can be expressed by the T-fraction

$$\sum_{n=0}^{\infty} \bar{W}_n(x) t^n = \frac{1}{1 - \frac{t}{1 - xt - \frac{2t}{1 - 2xt - \frac{3t}{1 - \dots}}}} \quad (7.9)$$

with coefficients  $\delta_i = (i - 1)x$  and  $\alpha_i = i$ .

For the special case  $x = 1$ , this T-fraction was known previously, at least empirically [117].

As preparation for this proof, let us mention two different combinatorial interpretations of the Ward numbers. The first one, which is well known, is in terms of phylogenetic trees<sup>1</sup>, the study of which goes back to Schröder's

---

<sup>1</sup> Also called labelled hierarchies [144, 107] [106, pp. 128–129 and 472–474], Schröder systems [181, 68] [69, pp. 223–224] or total partitions [193, Example 5.2.5 and Exercise 5.40].



Figure 7.1: (a) A wiggly line in an augmented perfect matching. (b) The two possibilities for a dashed line in a super-augmented perfect matching.

fourth problem [181]. A *phylogenetic tree of type  $(n, k)$*  is a rooted tree that has  $n + 1$  labelled leaves (numbered  $1, \dots, n + 1$ ) and  $k$  unlabelled internal vertices, in which every internal vertex has at least two children. It is easy to see that we must have  $1 \leq k \leq n$ , and not difficult to see that for  $n \geq 1$  the number of phylogenetic trees of type  $(n, k)$  is precisely the Ward number  $W(n, k)$ .<sup>2</sup> [*Proof.* The claim is clear for  $n = 1$ ; now we use induction. Leaf number  $n + 1$  can be connected either to an internal vertex on a phylogenetic tree of type  $(n - 1, k)$ , hence in  $kW(n - 1, k)$  ways; or to an edge on a phylogenetic tree of type  $(n - 1, k - 1)$ , creating a new internal vertex of out-degree 2, hence in  $(n + k - 2)W(n - 1, k - 1)$  ways; or to a new root, having the old root of a phylogenetic tree of type  $(n - 1, k - 1)$  as its other child, hence in  $W(n - 1, k - 1)$  ways. This proves (7.6).] So for  $n \geq 1$  the Ward polynomial  $W_n(x)$  is the generating polynomial for phylogenetic trees on  $n + 1$  labelled leaves in which each internal vertex gets a weight  $x$ .

The second interpretation, which seems to be new, is in terms of what we shall call *augmented perfect matchings*. Recall that a perfect matching of  $[2n] \stackrel{\text{def}}{=} \{1, \dots, 2n\}$  is simply a partition of  $[2n]$  into  $n$  pairs. We call the smaller (resp. larger) element of each pair an opener (resp. closer). An augmented perfect matching is then a perfect matching in which we may optionally also draw a wiggly line on an edge  $(i, i + 1)$  whenever  $i$  is a closer and  $i + 1$  is an opener [see Figure 7.1(a)]. Then the number of augmented perfect matchings of  $[2n]$  with  $n - k$  wiggly lines (or equivalently,  $k$  closers that are not followed by a wiggly line) is precisely the Ward number  $W(n, k)$ . [*Proof.* Let  $M(n, \ell)$  be the the number of augmented perfect matchings of  $[2n]$  with  $\ell$  wiggly lines. The number of ways that the vertex  $2n$  can be

<sup>2</sup> For  $n = 0$  there is a slight discrepancy: there are no phylogenetic trees with only one leaf, but we have defined  $W(0, k) = \delta_{0k}$ , not  $W(0, k) = 0$ .

Note also that the case  $k = n$  corresponds to binary phylogenetic trees, i.e. those in which every internal vertex has exactly two children. The number of such trees is  $W(n, n) = (2n - 1)!!$  [193, Example 5.2.6] [58, Section 2.5] [75, 172].

connected to a vertex  $i \in [2n - 1]$  that is not incident on a wiggly line is  $(2n - 1 - \ell)M(n - 1, \ell)$ , because the vertex  $i$  can be inserted in any of the gaps of an augmented perfect matching of  $[2n - 2]$  with  $\ell$  wiggly lines, where “gap” means a space between adjacent vertices that are not joined by a wiggly line, or a space at the start or end. The number of ways that the vertex  $2n$  can be connected to a vertex  $i \in [2n - 1]$  that is incident on a wiggly line is  $(n - \ell)M(n - 1, \ell - 1)$ , because the vertex  $i$  together with its incident wiggly line (pointing to the left) can be inserted after any closer of an augmented perfect matching of  $[2n - 2]$  with  $\ell - 1$  wiggly lines that is not incident on a wiggly line. Comparing the recurrence  $M(n, \ell) = (2n - 1 - \ell)M(n - 1, \ell) + (n - \ell)M(n - 1, \ell - 1)$  with (7.6), we see that  $M(n, \ell) = W(n, n - \ell)$ .] So the reversed Ward polynomial  $\overline{W}_n(x)$  is the generating polynomial for augmented perfect matchings of  $[2n]$  in which each wiggly line gets a weight  $x$ .

In Section 7.5 we construct a bijection between phylogenetic trees of type  $(n, k)$  and augmented perfect matchings of  $[2n]$  with  $n - k$  wiggly lines.

But we can go further: let us define a *super-augmented perfect matching* of  $[2n]$  to be a perfect matching of  $[2n]$  in which we may optionally draw a wiggly line on an edge  $(i, i + 1)$  whenever  $i$  is a closer and  $i + 1$  is an opener, and also optionally draw a dashed line on an edge  $(i, i + 1)$  whenever  $i$  is an opener and  $i + 1$  is a closer; however, it is not allowed for a wiggly line and a dashed line to be incident on the same vertex. Note that for a pair  $(i, i + 1)$  connected by a dashed line, there are two possibilities [see Figure 7.1(b)]: the opener  $i$  and the closer  $i + 1$  could belong to different arches (which necessarily cross), or they could belong to the same arch.

Now let  $M(n, \ell, m)$  be the number of super-augmented perfect matchings of  $[2n]$  with  $\ell$  wiggly lines and  $m$  dashed lines, and define the generating polynomials

$$\overline{W}_n(x, y, z) = \sum_{\ell, m} M(n, \ell, m) x^\ell y^m z^{n - \ell - m} . \quad (7.10)$$

(The variable  $z$  is of course redundant, but it is convenient to introduce it because it makes the polynomial homogeneous of degree  $n$ .) From what has already been said, we know that  $\overline{W}_n(x, 0, 1) = \overline{W}_n(x)$ . These three-variable polynomials have a T-fraction that generalises (7.9):

**Theorem 7.2.** *The ordinary generating function of the polynomials  $\overline{W}_n(x, y, z)$*

can be expressed by the T-fraction

$$\sum_{n=0}^{\infty} \overline{W}_n(x, y, z) t^n = \frac{1}{1 - yt - \frac{zt}{1 - (x + 2y)t - \frac{2zt}{1 - (2x + 3y)t - \frac{3zt}{1 - \dots}}}} \quad (7.11)$$

with coefficients  $\delta_i = (i - 1)x + iy$  and  $\alpha_i = iz$ .

Let us remark that  $\overline{W}_n(0, y, 1)$  is [198, reversal of A112493], while  $\overline{W}_n(1 - y, y, 1)$  is [198, A298673].

In Section 7.6 we will show that the polynomials  $\overline{W}_n(x, y, z)$  satisfy a linear differential recurrence, and that a sequence of four-variable polynomials generalizing  $\overline{W}_n(x, y, z)$  satisfies a non-linear differential recurrence.

But we can go much further: the main result of this chapter is a “master T-fraction” that enumerates super-augmented perfect matchings with an *infinite* set of independent statistics (Theorem 7.3). This master T-fraction includes many combinatorially interesting polynomials as special cases: see Corollary 7.4 and the specializations (7.48)–(7.51) and (7.52)–(7.54). By further specialization of (7.52)–(7.54) we obtain Theorems 7.2 and 7.1.

Here is one special case: Consider the linear recurrence

$$\widehat{W}(n, k) = (k + y) \widehat{W}(n - 1, k) + (n + k - 1) \widehat{W}(n - 1, k - 1) \quad (7.12)$$

where  $y$  is an indeterminate, and define the row-generating polynomials

$$W_n(x, y) = \sum_{k=0}^n \widehat{W}(n, k) x^k. \quad (7.13)$$

Then  $W_n(x, y)$  is the generating polynomial for super-augmented perfect matchings with a weight  $x$  for every closer that is not attached to a wiggly line, a weight  $y$  for every nearest-neighbour pair that is connected both by an ordinary line and by a dashed line, and a weight 0 for all other types of wiggly or dashed lines. Moreover, this will give a T-fraction with  $\alpha_i = ix$  and  $\delta_i = y + i - 1$  — which we might want to homogenize by writing  $\delta_i = y + (i - 1)z$ . It seems to be non-trivial that these polynomials satisfy a simple linear recurrence.

Recall that we considered another generalisation of perfect matchings in Section 2.2, namely *coloured* perfect matchings, where crossing arches are forbidden from sharing the same colour. We also considered the related problem of enumerating  $m$ -colourable perfect matchings (that is, perfect matching

whose edges can be  $m$ -coloured in such a way that crossing edges are assigned different colours), and showed that these are in bijection with permutations sortable by  $m$ -stacks in parallel. This problems appear to be more difficult, and we have not found continued fractions for their enumeration except in the trivial case  $m = 1$ , where these perfect matchings are enumerated by Catalan numbers.

Let us conclude this introduction by briefly mentioning some connections of the Ward numbers and Ward polynomials with other combinatorial objects and problems:

1) The *2-associated Stirling subset number*  $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}_{\geq 2}$  is the number of partitions of an  $n$ -element set into  $k$  blocks, each of which has at least two elements; by convention we set  $\left\{ \begin{smallmatrix} 0 \\ k \end{smallmatrix} \right\}_{\geq 2} = \delta_{0k}$ . It is easy to see that these numbers satisfy the recurrence

$$\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}_{\geq 2} = k \left\{ \begin{smallmatrix} n-1 \\ k \end{smallmatrix} \right\}_{\geq 2} + (n-1) \left\{ \begin{smallmatrix} n-2 \\ k-1 \end{smallmatrix} \right\}_{\geq 2} \quad \text{for } n \geq 1. \quad (7.14)$$

The first few  $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}_{\geq 2}$  are [69, pp. 221–222] [198, A008299/A137375]

$n \setminus k$	0	1	2	3	4	5	6	7	8	9	10	Row sums
0	1											1
1	0	0										0
2	0	1	0									1
3	0	1	0	0								1
4	0	1	3	0	0							4
5	0	1	10	0	0	0						11
6	0	1	25	15	0	0	0					41
7	0	1	56	105	0	0	0	0				162
8	0	1	119	490	105	0	0	0	0			715
9	0	1	246	1918	1260	0	0	0	0	0		3425
10	0	1	501	6825	9450	945	0	0	0	0	0	17722

Comparing (7.6) with (7.14), it is easy to see that  $W(n, k) = \left\{ \begin{smallmatrix} n+k \\ k \end{smallmatrix} \right\}_{\geq 2}$  [68] [62, eqns. (1.7) and (3.6)] [177, pp. 6–7]. This also has a simple bijective proof [68]: given a phylogenetic tree of type  $(n, k)$ , partition its  $n + k$  non-root vertices into blocks consisting of the sets of children of each of the  $k$  internal vertices.<sup>3</sup>

In fact, this argument gives more. Let  $I$  be any subset of  $\mathbb{N}$ ; let  $W_I(n, k)$  be the number of rooted trees on  $n + 1$  labelled leaves and  $k$  unlabelled

<sup>3</sup> Here the internal vertices are *a priori* unlabelled; but each internal vertex can be unambiguously labelled by the set of labels of its descendant leaves. So this is indeed a well-defined partition of  $n + k$  distinguishable objects.

internal vertices, in which the number of children of each internal vertex belongs to the set  $I$ ; and let  $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}_{\in I}$  be the number of partitions of an  $n$ -element set into  $k$  blocks, each of which has size belonging to the set  $I$ . Then  $W_I(n, k) = \left\{ \begin{smallmatrix} n+k \\ k \end{smallmatrix} \right\}_{\in I}$ .

2) A *Stirling permutation* [116] of order  $n$  is a permutation  $\sigma_1 \cdots \sigma_{2n}$  of the multiset  $\{1, 1, 2, 2, \dots, n, n\}$  in which, for all  $m \in [n]$ , all numbers between the two occurrences of  $m$  are larger than  $m$ . We say that an index  $j \in [2n]$  is a *descent* if  $\sigma_j > \sigma_{j+1}$  or  $j = 2n$  (so by our definition the last index is always a descent). The *second-order Eulerian number*  $\left\langle \begin{smallmatrix} n \\ k \end{smallmatrix} \right\rangle$  is the number of Stirling permutations of order  $n$  with exactly  $k$  descents; by convention we set  $\left\langle \begin{smallmatrix} 0 \\ k \end{smallmatrix} \right\rangle = \delta_{0k}$ .<sup>4</sup> It is not difficult to see [116, p. 27] that these numbers satisfy the recurrence

$$\left\langle \begin{smallmatrix} n \\ k \end{smallmatrix} \right\rangle = k \left\langle \begin{smallmatrix} n-1 \\ k \end{smallmatrix} \right\rangle + (2n-k) \left\langle \begin{smallmatrix} n-1 \\ k-1 \end{smallmatrix} \right\rangle \quad \text{for } n \geq 1. \quad (7.15)$$

The first few  $\left\langle \begin{smallmatrix} n \\ k \end{smallmatrix} \right\rangle$  are [198, A008517/A201637/A112007/A288874]

$n \setminus k$	0	1	2	3	4	5	6	7	8	Row sums
0	1									1
1	0	1								1
2	0	1	2							3
3	0	1	8	6						15
4	0	1	22	58	24					105
5	0	1	52	328	444	120				945
6	0	1	114	1452	4400	3708	720			10395
7	0	1	240	5610	32120	58140	33984	5040		135135
8	0	1	494	19950	195800	644020	785304	341136	40320	2027025

It is easy to see from (7.15) that  $\left\langle \begin{smallmatrix} n \\ n \end{smallmatrix} \right\rangle = n!$  and  $\sum_{k=0}^n \left\langle \begin{smallmatrix} n \\ k \end{smallmatrix} \right\rangle = (2n-1)!!$ . Define now the *second-order Eulerian polynomials*

$$E_n^{[2]}(x) = \sum_{k=0}^n \left\langle \begin{smallmatrix} n \\ k \end{smallmatrix} \right\rangle x^k \quad (7.16)$$

and the reversed polynomials  $\overline{E}_n^{[2]}(x) = x^n E_n^{[2]}(1/x)$ ; note that  $\overline{E}_n^{[2]}(0) = n!$  and  $\overline{E}_n^{[2]}(1) = (2n-1)!!$ . By translating the recurrences (7.6) and (7.15) into differential recurrences for the polynomials  $\overline{W}_n(x)$  and  $\overline{E}_n^{[2]}(x)$ , it is not difficult to show that

$$\overline{W}_n(x) = \overline{E}_n^{[2]}(1+x) \quad (7.17)$$

---

<sup>4</sup> **Warning:** Our definition of the second-order Eulerian numbers agrees with Gessel and Stanley [116] (who use the notation  $B_{n,k}$ ) but disagrees with Graham, Knuth and Patashnik [118, p. 270] because we define the last index to be a descent while they do not: thus  $\left\langle \begin{smallmatrix} 0 \\ k \end{smallmatrix} \right\rangle_{\text{ours}} = \left\langle \begin{smallmatrix} 0 \\ k \end{smallmatrix} \right\rangle^{\text{GKP}} = \delta_{0k}$  but  $\left\langle \begin{smallmatrix} n \\ k \end{smallmatrix} \right\rangle_{\text{ours}} = \left\langle \begin{smallmatrix} n \\ k-1 \end{smallmatrix} \right\rangle^{\text{GKP}}$  for  $n \geq 1$ .



or equivalently

$$E_n^{[2]}(x) = (1-x)^n W_n\left(\frac{x}{1-x}\right) \quad (7.18a)$$

$$W_n(x) = (1+x)^n E_n^{[2]}\left(\frac{x}{1+x}\right) \quad (7.18b)$$

In particular we have  $\overline{W}_n(-1) = n!$  [59, Theorem 3.1 and Section 4].

The identity (7.17) also has a nice combinatorial interpretation and proof. In a perfect matching of  $[2n]$ , let us say that a pair  $(i, i+1)$  is a closer/opener pair if  $i$  is a closer and  $i+1$  is an opener. Note that counting augmented perfect matchings with a weight  $x$  for each wiggly line — as  $\overline{W}_n(x)$  does — is equivalent to counting perfect matchings with a weight  $1+x$  for each closer/opener pair. Now let  $M'(n, \ell)$  be the number of perfect matchings of  $[2n]$  with  $\ell$  closer/opener pairs. We claim that  $M'(n, \ell) = \langle\langle n \rangle\rangle_{n-\ell}$ . [*Proof.* Write  $\text{clop}(\pi)$  for the number of closer/opener pairs in a perfect matching  $\pi$ . Now consider a perfect matching  $\pi$  of  $[2n]$ , and suppose that the vertex  $2n$  is paired with  $i \in [2n-1]$ ; let  $\pi'$  be the perfect matching of  $[2n-2]$  obtained from  $\pi$  by deleting the arch  $(i, 2n)$  and renumbering vertices. If  $i = 1$  or  $i - 1$  is an opener, then  $\text{clop}(\pi') = \text{clop}(\pi)$ . If  $i \geq 2$  and  $i - 1$  is a closer, then  $(i - 1, i)$  is a closer/opener pair; if also  $i + 1$  is an opener, then  $\text{clop}(\pi') = \text{clop}(\pi)$ , otherwise  $\text{clop}(\pi') = \text{clop}(\pi) - 1$ . If  $\text{clop}(\pi) = \ell$ , these insertions can be done in  $n, \ell$  and  $n - \ell$  ways, respectively. So  $M'(n, \ell) = (n + \ell)M'(n - 1, \ell) + (n - \ell)M'(n - 1, \ell - 1)$ , which is equivalent to (7.15).]<sup>5</sup> So the reversed second-order Eulerian polynomial  $\overline{E}_n^{[2]}(x)$  is the generating polynomial for perfect matchings of  $[2n]$  in which each closer/opener pair gets a weight  $x$ . This interpretation seems to be new.

3) There is also an interesting multivariate generalisation of the Ward polynomials. As explained above, the Ward polynomial  $W_n(x)$  is the generating polynomial for phylogenetic trees on  $n + 1$  labelled leaves in which each internal vertex gets a weight  $x$ . Now let  $\mathbf{x} = (x_i)_{i \geq 1}$  be an infinite collection of indeterminates, and let  $W_n(\mathbf{x}) = W_n(x_1, \dots, x_n)$  be the generating polynomial for phylogenetic trees on  $n + 1$  labelled leaves in which each internal vertex with  $i$  ( $\geq 2$ ) children gets a weight  $x_{i-1}$ ; we call this the *multivariate Ward polynomial*. A straightforward argument (generalizing [193, Examples 5.2.5 and 5.2.6] or [106, pp. 128–129]) shows that the exponential

<sup>5</sup> Note also that  $M'(n, 0) = n!$  has an easy proof: if there are no closer/opener pairs, then the elements  $1, \dots, n$  are openers and  $n+1, \dots, 2n$  are closers, and a perfect matching is obtained by pairing these elements in one of the  $n!$  possible ways.

generating function

$$\mathcal{W}(t; \mathbf{x}) \stackrel{\text{def}}{=} \sum_{n=1}^{\infty} W_n(\mathbf{x}) \frac{t^n}{n!} \quad (7.19)$$

(note that here the sum starts at  $n = 1$ ) satisfies the functional equation

$$\mathcal{W}(t; \mathbf{x}) = t + \sum_{n=2}^{\infty} x_{n-1} \frac{\mathcal{W}(t; \mathbf{x})^n}{n!} . \quad (7.20)$$

In other words,  $\mathcal{W}(t; \mathbf{x})$  is the compositional inverse of the generic power series

$$F(t; \mathbf{x}) \stackrel{\text{def}}{=} t - \sum_{n=2}^{\infty} x_{n-1} \frac{t^n}{n!} . \quad (7.21)$$

In this latter context the polynomials  $W_n(x_1, \dots, x_n)$  can be found in the books of Riordan [176, p. 181]<sup>6</sup> and Comtet [69, p. 151], but without the interpretation in terms of phylogenetic trees. It would be interesting to extend our work on the Ward polynomials to this multivariate generalisation. Unfortunately, we do not know any simple specialisations of  $W_n(\mathbf{x})$ , other than  $x_i = x$  for all  $i$  that also yield a simple T-fraction for their ordinary generating function. We do know a fairly complicated one, where  $F(t; \mathbf{x}) = (1+x)/w \log(1+wt) - x[(1+wt)^{1/w} - 1]$ , corresponding to  $\delta_i = i - 1 + w$  and  $\alpha_i = ix$ . It follows that the exponential generating function corresponding to this T-fraction as ogf, call it  $\mathcal{W}(t; x, w)$ , satisfies  $\mathcal{W}(t; x, w) = w^{-1} \log[1 + w\mathcal{W}(t; x, 0)]$ . Here  $\mathcal{W}(t; x, 0)$  is the egf of the Ward polynomials, which is the compositional inverse of  $F(t; x) = t - x(e^t - 1 - t)$ .

4) Finally, let us comment on the implications of our results for the theory of Hankel-total positivity [186, 188], which was in fact our original motivation for studying these T-fractions. If  $\mathbf{P} = (P_n(\mathbf{x}))_{n \geq 0}$  is a sequence of polynomials with real coefficients in one or more indeterminates  $\mathbf{x}$ , we say that this sequence is *coefficientwise Hankel-totally positive* if every minor of the infinite Hankel matrix  $H_{\infty}(\mathbf{P}) = (P_{i+j}(\mathbf{x}))_{i,j \geq 0}$  is a polynomial with nonnegative coefficients. One fundamental result of this theory is [188, 166]: whenever the ordinary generating function  $\sum_{n=0}^{\infty} P_n(\mathbf{x}) t^n$  is given by a T-fraction (7.3) where all the coefficients  $\alpha_i$  and  $\delta_i$  are polynomials in  $\mathbf{x}$  with nonnegative

---

<sup>6</sup> With the erratum given by Riordan in [177, p. 7]. Riordan [177, p. 7] also observes (attributing this remark to Neil Sloane) that the specialisation of  $W_n(x_1, \dots, x_n)$  [defined via compositional inverses] to  $x_1 = \dots = x_n = x$  enumerates phylogenetic trees by number of internal vertices. But his formula  $S_n(y) = Z_n(y, \dots, y)$  [his  $Z_n$  is our  $W_n$ ] should read  $S_n(y) = Z_{n-1}(y, \dots, y)$ , since his  $S_n(y)$  enumerates phylogenetic trees on  $n$  leaves, not  $n + 1$ .

coefficients, the sequence  $\mathbf{P} = (P_n(\mathbf{x}))_{n \geq 0}$  is coefficientwise Hankel-totally positive<sup>7</sup>. It follows from this that almost all of the T-fractions derived in this chapter — including the most general one, Theorem 7.3 — are coefficientwise Hankel-totally positive in the appropriate variables. In particular, the Ward polynomials  $W_n(x)$  and the generalised reversed Ward polynomials  $\overline{W}_n(x, y, z)$  are coefficientwise Hankel-totally positive.

It follows from (7.17) and (7.9) that the ordinary generating function of the reversed second-order Eulerian polynomials is given by a T-fraction with coefficients  $\delta_i = (i - 1)(x - 1)$  and  $\alpha_i = i$ .<sup>8</sup> Here  $\delta_i$  is *not* coefficientwise nonnegative (or even pointwise nonnegative when  $0 \leq x < 1$ ), so the general theory [188, 166] says nothing about the Hankel-total positivity of the reversed second-order Eulerian polynomials. And yet, we find empirically that the sequence of reversed second-order Eulerian polynomials *is* coefficientwise Hankel-totally positive: we have tested this through the  $13 \times 13$  Hankel matrix. This mystery has now been resolved [166]. The reversed second-order Eulerian polynomials  $\overline{E}_n^{[2]}(x)$  are also given by a 2-branched S-fraction [166] with coefficients  $\boldsymbol{\alpha} = (\alpha_i)_{i \geq 2} = 1, 1, x, 2, 2, 2x, 3, 3, 3x, \dots$ . Since these  $\alpha_i$  are coefficientwise nonnegative, the general theory of branched S-fractions [166] implies the coefficientwise Hankel-total positivity.

The plan of this chapter is as follows: In Section 7.2 we state our results on the enumeration of super-augmented perfect matchings: these include a very general T-fraction with infinitely many indeterminates (Theorem 7.3) as well as numerous special cases that count statistics of combinatorial interest. In Section 7.3 we recall some basic facts concerning the combinatorial interpretation of continued fractions and the concept of labelled Schröder paths. In Section 7.4 we prove Theorem 7.3 by constructing a bijection from the set of super-augmented perfect matchings onto a set of labelled 2-coloured Schröder paths. In Section 7.5 we give a bijective proof of a relationship between the enumeration of augmented perfect matchings and phylogenetic trees. Finally, in Section 7.6, we use Theorem 7.3 to show that the series defined as a general linear T-fraction, defined by  $\alpha_i = x + (i - 1)u$  and  $\delta_i = z + (i - 1)w$ , satisfies a differential recurrence. As a corollary, we show that this series is D-algebraic as a series in  $t$  for any fixed values of  $x$ ,  $u$ ,  $z$  and  $w$  satisfying  $u + w = 0$ .

---

<sup>7</sup>This is generalised in Section 9 of [166]

<sup>8</sup> This T-fraction is also a consequence of [58, Section 5.2, item (7)] together with the Roblet–Viennot [178] interpretation of T-fractions as a generating function for Dyck paths with special weights for peaks.

## Statement of results

We begin by reviewing some S-fractions for perfect matchings that were recently obtained by Sokal and Zeng [189]. We then state our generalisations, which are T-fractions for super-augmented perfect matchings.

### S-fractions for perfect matchings

Euler showed [98, section 29] that the generating function of the odd semi-factorials can be represented as an S-type continued fraction

$$\sum_{n=0}^{\infty} (2n-1)!! t^n = \frac{1}{1 - \frac{t}{1 - \frac{2t}{1 - \frac{3t}{1 - \dots}}}} \quad (7.22)$$

with coefficients  $\alpha_n = n$ .<sup>9</sup> Since  $(2n-1)!!$  enumerates perfect matchings of a  $2n$ -element set, it is natural to seek polynomial refinements of this sequence that enumerate perfect matchings of  $[2n]$  according to some natural statistic(s). Note that we can regard a perfect matching either as a special type of set partition (namely, one in which all blocks are of size 2) or as a special type of permutation (namely, a fixed-point-free involution). We will adopt here the former interpretation, and write  $\pi \in \mathcal{M}_{2n} \subseteq \Pi_{2n}$ . If  $i, j \in [2n]$  are paired in the perfect matching  $\pi$ , we write  $i \sim_{\pi} j$  (or just  $i \sim j$  if  $\pi$  is clear from the context).

Inspired by (7.22), let us introduce the polynomials  $M_n(x, y, u, v)$  defined by the continued fraction

$$\sum_{n=0}^{\infty} M_n(x, y, u, v) t^n = \frac{1}{1 - \frac{xt}{1 - \frac{(y+v)t}{1 - \frac{(x+2u)t}{1 - \frac{(y+3v)t}{1 - \frac{(x+4u)t}{1 - \frac{(y+5v)t}{1 - \dots}}}}} \quad (7.23)$$

---

<sup>9</sup> See also [58, Section 2.6] for a combinatorial proof of (7.22) based on a counting of height-labelled Dyck paths.

with coefficients

$$\alpha_{2k-1} = x + (2k-2)u \quad (7.24a)$$

$$\alpha_{2k} = y + (2k-1)v \quad (7.24b)$$

Clearly  $M_n(x, y, u, v)$  is a homogeneous polynomial of degree  $n$ . Since  $M_n(1, 1, 1, 1) = (2n-1)!!$ , it is natural to expect that  $M_n(x, y, u, v)$  enumerates perfect matchings of  $[2n]$  according to some natural trivariate statistic. Let us now explain, following [189], what this statistic is.

Let  $\pi$  be a perfect matching of  $[2n]$ . We recall that a vertex  $i \in [2n]$  is called an opener (resp. closer) if it is the smaller (resp. larger) element of its pair. Let us say that an opener  $j$  (with corresponding closer  $k$ ) is a *record* if there does not exist an opener  $i < j$  that is paired with a closer  $l > k$ . In other words,  $j$  is a record if there does not exist an arch  $(i, l)$  that nests above the arch  $(j, k)$ . Similarly, let us say that a closer  $k$  (with corresponding opener  $j$ ) is an *antirecord* if there does not exist a closer  $l > k$  that is paired with an opener  $i < j$ . In other words,  $k$  is an antirecord if and only if  $j$  is a record.<sup>10</sup>

Let us now classify the closers of  $\pi$  into four types:

- *even closer antirecord* (ecar) [i.e.  $i$  is even and is an antirecord];
- *odd closer antirecord* (ocar) [i.e.  $i$  is odd and is an antirecord];
- *even closer non-antirecord* (ecnar) [i.e.  $i$  is even and is not an antirecord];
- *odd closer non-antirecord* (ocnar) [i.e.  $i$  is odd and is not an antirecord].

Similarly, we classify the openers of  $\pi$  into four types:

- *even opener record* (eor) [i.e.  $i$  is even and is a record];
- *odd opener record* (oor) [i.e.  $i$  is odd and is a record];
- *even opener non-record* (eonr) [i.e.  $i$  is even and is not a record];
- *odd opener non-record* (oonr) [i.e.  $i$  is odd and is not a record].

---

<sup>10</sup> This terminology of records and antirecords comes from the alternate interpretation of perfect matchings as fixed-point-free involutions. In general, for a permutation  $\sigma \in \mathfrak{S}_n$ , an index  $i \in [n]$  is called a *record* (or *left-to-right maximum*) if  $\sigma(j) < \sigma(i)$  for all  $j < i$ , and is called an *antirecord* (or *right-to-left minimum*) if  $\sigma(j) > \sigma(i)$  for all  $j > i$ . See [189] for further discussion of record-antirecord statistics in permutations.

Then Sokal and Zeng [189] proved that the polynomials  $M_n(x, y, u, v)$  defined by (7.23)/(7.24) have the combinatorial interpretation

$$M_n(x, y, u, v) = \sum_{\pi \in \mathcal{M}_{2n}} x^{\text{ecar}(\pi)} y^{\text{ocar}(\pi)} u^{\text{ecnar}(\pi)} v^{\text{ocnar}(\pi)} \quad (7.25a)$$

$$= \sum_{\pi \in \mathcal{M}_{2n}} x^{\text{oor}(\pi)} y^{\text{eor}(\pi)} u^{\text{oonr}(\pi)} v^{\text{eonr}(\pi)} . \quad (7.25b)$$

The interpretations (7.25a) and (7.25b) are of course trivially equivalent under the involution  $i \rightarrow 2n + 1 - i$ , which preserves the structure of a perfect matching but interchanges even with odd, opener with closer, and record with antirecord.

Sokal and Zeng [189] also generalised the four-variable polynomial  $M_n(x, y, u, v)$  by adding weights for crossings and nestings. Given a perfect matching  $\pi \in \mathcal{M}_{2n}$ , we say that a quadruplet  $i < j < k < l$  forms a *crossing* if  $i \sim_\pi k$  and  $j \sim_\pi l$ , and a *nesting* if  $i \sim_\pi l$  and  $j \sim_\pi k$ . We write  $\text{cr}(\pi)$  and  $\text{ne}(\pi)$  for the numbers of crossings and nestings of  $\pi$ . We now define the six-variable polynomial

$$M_n(x, y, u, v, p, q) = \sum_{\pi \in \mathcal{M}_{2n}} x^{\text{ecar}(\pi)} y^{\text{ocar}(\pi)} u^{\text{ecnar}(\pi)} v^{\text{ocnar}(\pi)} p^{\text{cr}(\pi)} q^{\text{ne}(\pi)} \quad (7.26a)$$

$$= \sum_{\pi \in \mathcal{M}_{2n}} x^{\text{oor}(\pi)} y^{\text{eor}(\pi)} u^{\text{oonr}(\pi)} v^{\text{eonr}(\pi)} p^{\text{cr}(\pi)} q^{\text{ne}(\pi)} . \quad (7.26b)$$

Here the equality of (7.26a) and (7.26b) follows again from the involution  $i \rightarrow 2n + 1 - i$ , which preserves the numbers of crossings and nestings. Sokal and Zeng [189] showed that the ordinary generating function of the polynomials  $M_n(x, y, u, v, p, q)$  has an S-fraction that generalises (7.23)/(7.24), namely

$$\sum_{n=0}^{\infty} M_n(x, y, u, v, p, q) t^n = \frac{1}{1 - \frac{xt}{1 - \frac{(py + qv)t}{1 - \frac{(p^2x + q[2]_{p,q}u)t}{1 - \frac{(p^3y + q[3]_{p,q}v)t}{1 - \dots}}}}} \quad (7.27)$$

with coefficients

$$\alpha_{2k-1} = p^{2k-2}x + q[2k-2]_{p,q}u \quad (7.28a)$$

$$\alpha_{2k} = p^{2k-1}y + q[2k-1]_{p,q}v \quad (7.28b)$$

where

$$[n]_{p,q} \stackrel{\text{def}}{=} \frac{p^n - q^n}{p - q} = \sum_{j=0}^{n-1} p^j q^{n-1-j} \quad (7.29)$$

for an integer  $n \geq 0$ . When  $p = q = 1$  this reduces to (7.23)/(7.24). Note that if  $u = x$  and/or  $v = y$ , then the weights (7.28) simplify to  $\alpha_{2k-1} = [2k-1]_{p,q} x$  and  $\alpha_{2k} = [2k]_{p,q} y$ , respectively. For the special case  $x = y = u = v$ , the S-fraction (7.27) was obtained previously by Kasraoui and Zeng [134] (see also [31, p. 3280]).

But Sokal and Zeng [189] went even further, by defining statistics that count the number of quadruplets  $i < j < k < l$  that form crossings or nestings *with a particular vertex  $k$  in third position*:

$$\text{cr}(k, \pi) = \#\{i < j < k < l: i \sim_{\pi} k \text{ and } j \sim_{\pi} l\} \quad (7.30)$$

$$\text{ne}(k, \pi) = \#\{i < j < k < l: i \sim_{\pi} l \text{ and } j \sim_{\pi} k\} \quad (7.31)$$

Note that  $\text{cr}(k, \pi)$  and  $\text{ne}(k, \pi)$  can be nonzero only when  $k$  is a closer (and  $k \geq 3$ ); and we obviously have

$$\text{cr}(\pi) = \sum_{k \in \text{closers}} \text{cr}(k, \pi) \quad (7.32a)$$

$$\text{ne}(\pi) = \sum_{k \in \text{closers}} \text{ne}(k, \pi) \quad (7.32b)$$

Sokal and Zeng [189] also defined

$$\text{qne}(k, \pi) = \#\{i < k < l: i \sim_{\pi} l\} \quad (7.33)$$

(they call this a *quasi-nesting* of the vertex  $k$ ). We will use this statistic only when  $k$  is an opener: in this case  $\text{qne}(k, \pi)$  counts the number of times that the opener  $k$  occurs in *second* position in a crossing or nesting: when  $(i, l) \in \mathcal{G}_{\pi}$  is a pair contributing to  $\text{qne}(k, \pi)$ , the quadruplet  $i < k < l, m$  [where  $(k, m) \in \mathcal{G}_{\pi}$ ] must be either a crossing or nesting (according as  $m > l$  or  $m < l$ ), but we do not keep track of which one it is.

Now introduce two infinite families of indeterminates  $\mathbf{a} = (\mathbf{a}_{\ell})_{\ell \geq 0}$  and  $\mathbf{b} = (\mathbf{b}_{\ell, \ell'})_{\ell, \ell' \geq 0}$ , and define the polynomials  $M_n(\mathbf{a}, \mathbf{b})$  by

$$M_n(\mathbf{a}, \mathbf{b}) = \sum_{\pi \in \mathcal{M}_{2n}} \prod_{i \in \text{openers}} \mathbf{a}_{\text{qne}(i, \pi)} \prod_{i \in \text{closers}} \mathbf{b}_{\text{cr}(i, \pi), \text{ne}(i, \pi)} \cdot \quad (7.34)$$

Sokal and Zeng [189] showed that the ordinary generating function of the polynomials  $M_n(\mathbf{a}, \mathbf{b})$  has the S-fraction

$$\sum_{n=0}^{\infty} M_n(\mathbf{a}, \mathbf{b}) t^n = \frac{1}{1 - \frac{a_0 b_{00} t}{1 - \frac{a_1 (b_{01} + b_{10}) t}{1 - \frac{a_2 (b_{02} + b_{11} + b_{20}) t}{1 - \dots}}}} \quad (7.35)$$

with coefficients

$$\alpha_n = a_{n-1} b_{n-1}^* \quad (7.36)$$

where

$$b_{n-1}^* \stackrel{\text{def}}{=} \sum_{\ell=0}^{n-1} b_{\ell, n-1-\ell}. \quad (7.37)$$

This “master S-fraction for perfect matchings” contains (7.23)–(7.25) and (7.26)–(7.28) as special cases [189].

## T-fractions for super-augmented perfect matchings

Recall that a *super-augmented perfect matching* of  $[2n]$  is a perfect matching of  $[2n]$  in which we may optionally draw a wiggly line on an edge  $(i, i+1)$  whenever  $i$  is a closer and  $i+1$  is an opener, and also optionally draw a dashed line on an edge  $(i, i+1)$  whenever  $i$  is an opener and  $i+1$  is a closer; however, it is not allowed for a wiggly line and a dashed line to be incident on the same vertex. We denote a super-augmented perfect matching of  $[2n]$  by  $\tau \in \mathcal{M}_{2n}^*$ , and we write  $\pi(\tau) \in \mathcal{M}_{2n}$  for the underlying perfect matching. Of course, we say that  $i$  is an opener (resp. closer) in  $\tau$  in case it is an opener (resp. closer) in  $\pi(\tau)$ .

For  $\tau \in \mathcal{M}_{2n}^*$ , we call a vertex  $i$  a

- *pure opener* if it is an opener and not incident on any wiggly or dashed line;
- *pure closer* if it is a closer and not incident on any wiggly or dashed line.

We call a pair  $(i, i+1)$  a

- *wiggly pair* if  $i$  is a closer,  $i+1$  is an opener, and they are connected by a wiggly line;



- *dashed pair* if  $i$  is an opener,  $i + 1$  is a closer, and they are connected by a dashed line.

The statistics (7.30)–(7.33) apply equally well to super-augmented perfect matchings if we evaluate them on  $\pi = \pi(\tau)$ . We will generalise (7.34), not by introducing any new statistics, but simply by refining the distinctions among types of vertices. We introduce four infinite families of indeterminates  $\mathbf{a} = (a_\ell)_{\ell \geq 0}$ ,  $\mathbf{b} = (b_{\ell, \ell'})_{\ell, \ell' \geq 0}$ ,  $\mathbf{f} = (f_{\ell, \ell'})_{\ell, \ell' \geq 0}$  and  $\mathbf{g} = (g_{\ell, \ell'})_{\ell, \ell' \geq 0}$ , and define the polynomials  $M_n(\mathbf{a}, \mathbf{b}, \mathbf{f}, \mathbf{g})$  by

$$M_n(\mathbf{a}, \mathbf{b}, \mathbf{f}, \mathbf{g}) = \sum_{\pi \in \mathcal{M}_{2n}^*} \prod_{i \in \text{pureop}} a_{\text{qne}(i, \pi)} \prod_{i \in \text{purecl}} b_{\text{cr}(i, \pi), \text{ne}(i, \pi)} \prod_{(i, i+1) \in \text{wiggly}} f_{\text{cr}(i, \pi), \text{ne}(i, \pi)} \prod_{(i, i+1) \in \text{dashed}} g_{\text{cr}(i+1, \pi), \text{ne}(i+1, \pi)} \quad (7.38)$$

where pureop (resp. purecl) denotes the pure openers (resp. pure closers). Our main result is then the following:

**Theorem 7.3** (Master T-fraction for super-augmented perfect matchings). *The ordinary generating function of the polynomials  $M_n(\mathbf{a}, \mathbf{b}, \mathbf{f}, \mathbf{g})$  has the T-type continued fraction*

$$\sum_{n=0}^{\infty} M_n(\mathbf{a}, \mathbf{b}, \mathbf{f}, \mathbf{g}) t^n = \frac{1}{1 - g_{00}t - \frac{a_0 b_{00}t}{1 - (f_{00} + g_{01} + g_{10})t - \frac{a_1(b_{01} + b_{10})t}{1 - \dots}}} \quad (7.39)$$

with coefficients

$$\alpha_n = a_{n-1} b_{n-1}^* \quad (7.40a)$$

$$\delta_n = f_{n-2}^* + g_{n-1}^* \quad (7.40b)$$

where

$$b_{n-1}^* \stackrel{\text{def}}{=} \sum_{\ell=0}^{n-1} b_{\ell, n-1-\ell} \quad (7.41)$$

and likewise for  $\mathbf{f}$  and  $\mathbf{g}$ .

When  $\mathbf{f} = \mathbf{g} = 0$ , wiggly and dashed lines are forbidden, and Theorem 7.3 reduces to the S-fraction (7.35)/(7.36).

Let us now show some specialisations of Theorem 7.3 that generalise the S-fractions shown in the preceding subsection. We begin by defining a

14-variable polynomial  $M_n(x, y, u, v, x', y', u', v', x'', y'', u'', v'', p, q)$  that generalises (7.26a) by distinguishing three types of closers: pure closers, closers incident on a wiggly line, and closers incident on a dashed line. We let  $\text{ecar}^\circ(\tau)$ ,  $\text{ecar}'(\tau)$  and  $\text{ecar}''(\tau)$  be the number of even closer antirecords in each of these three classes, and likewise for the other statistics. We then define:

$$\begin{aligned} M_n(x, y, u, v, x', y', u', v', x'', y'', u'', v'', p, q) \\ = \sum_{\tau \in \mathcal{M}_{2n}^*} x^{\text{ecar}^\circ(\tau)} y^{\text{ocar}^\circ(\tau)} u^{\text{ecnar}^\circ(\tau)} v^{\text{ocnar}^\circ(\tau)} \\ (x')^{\text{ecar}'(\tau)} (y')^{\text{ocar}'(\tau)} (u')^{\text{ecnar}'(\tau)} (v')^{\text{ocnar}'(\tau)} \\ (x'')^{\text{ecar}''(\tau)} (y'')^{\text{ocar}''(\tau)} (u'')^{\text{ecnar}''(\tau)} (v'')^{\text{ocnar}''(\tau)} p^{\text{cr}(\tau)} q^{\text{ne}(\tau)} \end{aligned} \quad (7.42)$$

We have:

**Corollary 7.4.** *The ordinary generating function of the polynomials (7.42) has the T-type continued fraction*

$$\begin{aligned} \sum_{n=0}^{\infty} M_n(x, y, u, v, x', y', u', v', x'', y'', u'', v'', p, q) t^n \\ = \frac{1}{1 - x''t - \frac{xt}{1 - (x' + py'' + qv'')t - \frac{(py + qv)t}{1 - (py' + qv' + p^2x'' + q[2]_{p,1}u'')t - \frac{(p^2x + q[2]_{p,q}u)t}{1 - \dots}}}} \end{aligned} \quad (7.43)$$

with coefficients

$$\alpha_{2k-1} = p^{2k-2}x + q[2k-2]_{p,q}u \quad (7.44a)$$

$$\alpha_{2k} = p^{2k-1}y + q[2k-1]_{p,q}v \quad (7.44b)$$

$$\delta_1 = x'' \quad (7.44c)$$

$$\delta_{2k-1} = p^{2k-3}y' + q[2k-3]_{p,q}v' + p^{2k-2}x'' + q[2k-2]_{p,q}u'' \text{ for } k \geq 2 \quad (7.44d)$$

$$\delta_{2k} = p^{2k-2}x' + q[2k-2]_{p,q}u' + p^{2k-1}y'' + q[2k-1]_{p,q}v'' \quad (7.44e)$$

When  $x' = y' = u' = v' = x'' = y'' = u'' = v'' = 0$ , Corollary 7.4 reduces to the S-fraction (7.27)/(7.28).

The proof of Corollary 7.4 will be based on the following easy combinatorial lemma [189]:

**Lemma 7.5** (Closers in perfect matchings). *Let  $\pi \in \mathcal{M}_{2n}$  be a perfect matching of  $[2n]$ , and let  $k \in [2n]$  be a closer of  $\pi$ . Then:*

(a)  $k$  has the same parity as  $\text{cr}(k, \pi) + \text{ne}(k, \pi)$ .

(b)  $k$  is an antirecord if and only if  $\text{ne}(k, \pi) = 0$ .

*Proof.* (a) Let  $j$  be the opener that is paired with the closer  $k$ . For each  $i < k$ , let  $\sigma(i)$  ( $\neq i$ ) be the element with which it is paired. Then the set  $\{i: i < k\}$ , which has cardinality  $k - 1$ , can be partitioned as

$$\{j\} \cup \{i < k: \sigma(i) < k\} \cup \{i < j: \sigma(i) > k\} \cup \{j < i < k: \sigma(i) > k\}. \quad (7.45)$$

The first of these sets has cardinality 1; the second has even cardinality; the third has cardinality  $\text{ne}(k, \pi)$ ; and the fourth has cardinality  $\text{cr}(k, \pi)$ .

(b) As explained earlier, a closer  $k$  (with corresponding opener  $j$ ) is an antirecord if and only if there does not exist a closer  $l > k$  that is paired with an opener  $i < j$ . But this is precisely the statement that  $\text{ne}(k, \pi) = 0$ .  $\square$

*Proof of Corollary 7.4.* In (7.38) we set  $\mathbf{a}_\ell = 1$  for all  $\ell \geq 0$  and

$$\mathbf{b}_{\ell, \ell'} = p^\ell q^{\ell'} \times \begin{cases} x & \text{if } \ell' = 0 \text{ and } \ell \text{ is even} \\ y & \text{if } \ell' = 0 \text{ and } \ell \text{ is odd} \\ u & \text{if } \ell' \geq 1 \text{ and } \ell + \ell' \text{ is even} \\ v & \text{if } \ell' \geq 1 \text{ and } \ell + \ell' \text{ is odd} \end{cases} \quad (7.46a)$$

$$\mathbf{f}_{\ell, \ell'} = p^\ell q^{\ell'} \times \begin{cases} x' & \text{if } \ell' = 0 \text{ and } \ell \text{ is even} \\ y' & \text{if } \ell' = 0 \text{ and } \ell \text{ is odd} \\ u' & \text{if } \ell' \geq 1 \text{ and } \ell + \ell' \text{ is even} \\ v' & \text{if } \ell' \geq 1 \text{ and } \ell + \ell' \text{ is odd} \end{cases} \quad (7.46b)$$

$$\mathbf{g}_{\ell, \ell'} = p^\ell q^{\ell'} \times \begin{cases} x'' & \text{if } \ell' = 0 \text{ and } \ell \text{ is even} \\ y'' & \text{if } \ell' = 0 \text{ and } \ell \text{ is odd} \\ u'' & \text{if } \ell' \geq 1 \text{ and } \ell + \ell' \text{ is even} \\ v'' & \text{if } \ell' \geq 1 \text{ and } \ell + \ell' \text{ is odd} \end{cases} \quad (7.46c)$$

By Lemma 7.5(a,b) and (7.32a,b) we obtain the polynomial (7.42). Then

$$\mathbf{b}_{n-1}^\star \stackrel{\text{def}}{=} \sum_{\ell=0}^{n-1} \mathbf{b}_{\ell, n-1-\ell} = \begin{cases} p^{n-1}x + q[n-1]_{p,q}u & \text{if } n \text{ is odd} \\ p^{n-1}y + q[n-1]_{p,q}v & \text{if } n \text{ is even} \end{cases} \quad (7.47)$$

and similarly for  $\mathbf{f}$  and  $\mathbf{g}$ . The result then follows from Theorem 7.3.  $\square$

If we further specialise (7.42) by setting  $x = y$ ,  $u = v$ ,  $x' = y'$ ,  $u' = v'$ ,  $x'' = y''$ ,  $u'' = v''$ , we obtain 8-variable polynomials that count closers of each of three types (pure, wiggly or dashed) according to whether they are antirecords or not (but forgetting whether they are even or odd), and also count crossings and nestings:

$$M_n(x, x, u, u, x', x', u', u', x'', x'', u'', u'', p, q) = \sum_{\tau \in \mathcal{M}_{2n}^*} x^{\text{car}^\circ(\tau)} u^{\text{cnar}^\circ(\tau)} (x')^{\text{car}'(\tau)} (u')^{\text{cnar}'(\tau)} (x'')^{\text{car}''(\tau)} (u'')^{\text{cnar}''(\tau)} p^{\text{cr}(\tau)} q^{\text{ne}(\tau)} \quad (7.48)$$

Their T-fraction is

$$\sum_{n=0}^{\infty} M_n(x, x, u, u, x', x', u', u', x'', x'', u'', u'', p, q) t^n = \frac{1}{1 - x''t - \frac{xt}{1 - (x' + px'' + qu'')t - \frac{(px + qu)t}{1 - (px' + qu' + p^2x'' + q[2]_{p,1}u'')t - \frac{(p^2x + q[2]_{p,q}u)t}{1 - \dots}}}} \quad (7.49)$$

with coefficients

$$\alpha_n = p^{n-1}x + q[n-1]_{p,q}u \quad (7.50a)$$

$$\delta_1 = x'' \quad (7.50b)$$

$$\delta_n = p^{n-2}x' + q[n-2]_{p,q}u' + p^{n-1}x'' + q[n-1]_{p,q}u'' \text{ for } n \geq 2 \quad (7.50c)$$

Note that the forms of the coefficients (7.50) are no longer alternating between even and odd, because we are no longer distinguishing even closers from odd closers. If we further specialise this to  $p = q = 1$ , we get

$$\alpha_n = x + (n-1)u \quad (7.51a)$$

$$\delta_1 = x'' \quad (7.51b)$$

$$\delta_n = x' + (n-2)u' + x'' + (n-1)u'' \text{ for } n \geq 2 \quad (7.51c)$$

When  $u' = x'$  the formula (7.51c) holds also for  $n = 1$ , and the resulting polynomials  $M_n$  satisfy a recurrence (see Section 7.6).

On the other hand, if we specialise (7.42) by setting  $x = y = u = v = z$ ,  $x' = y' = u' = v' = \xi$  and  $x'' = y'' = u'' = v'' = \eta$ , we obtain 5-variable polynomials that count the numbers of wiggly and dashed lines as well as the numbers of crossings and nestings:

$$M_n(\xi, \eta, z, p, q) = \sum_{\tau \in \mathcal{M}_{2n}^*} \xi^{\text{wig}(\tau)} \eta^{\text{dash}(\tau)} z^{n - \text{wig}(\tau) - \text{dash}(\tau)} p^{\text{cr}(\tau)} q^{\text{ne}(\tau)}. \quad (7.52)$$

Their T-fraction is

$$\begin{aligned} & \sum_{n=0}^{\infty} M_n(\xi, \eta, z, p, q) t^n \\ &= \frac{1}{1 - \eta t - \frac{xt}{1 - (\xi + [2]_{p,q}\eta)t - \frac{[2]_{p,q}zt}{1 - ([2]_{p,q}\xi + [3]_{p,q}\eta)t - \frac{[3]_{p,q}zt}{1 - \dots}}}} \end{aligned} \quad (7.53)$$

with coefficients

$$\alpha_n = [n]_{p,q}z \quad (7.54a)$$

$$\delta_n = [n-1]_{p,q}\xi + [n]_{p,q}\eta \quad (7.54b)$$

Note again that the forms of the coefficients (7.54) are no longer alternating between even and odd.

If we further specialise (7.53)/(7.54) to  $p = q = 1$ , we obtain Theorem 7.2; and if we specialise that to  $\eta = 0$ , we obtain Theorem 7.1.

## Preliminaries

Our proof of Theorem 7.3 will be based on Flajolet’s [104] combinatorial interpretation of continued fractions in terms of Dyck and Motzkin paths, adapted slightly to handle Schröder paths, together with a bijection that maps super-augmented perfect matchings to labelled Schröder paths. We begin by briefly reviewing these two ingredients.

### Combinatorial interpretation of continued fractions

Recall that a *Motzkin path* of length  $n \geq 0$  is a path  $\omega = (\omega_0, \dots, \omega_n)$  in the right quadrant  $\mathbb{N}_0 \times \mathbb{N}_0$ , starting at  $\omega_0 = (0, 0)$  and ending at  $\omega_n = (n, 0)$ , whose steps  $s_j = \omega_j - \omega_{j-1}$  are  $(1, 1)$  [“rise”],  $(1, -1)$  [“fall”] or  $(1, 0)$  [“level”]. We write  $\mathcal{M}_n$  for the set of Motzkin paths of length  $n$ , and  $\mathcal{M} = \bigcup_{n=0}^{\infty} \mathcal{M}_n$ . A Motzkin path is called a *Dyck path* if it has no level steps. A Dyck path always has even length; we write  $\mathcal{D}_{2n}$  for the set of Dyck paths of length  $2n$ , and  $\mathcal{D} = \bigcup_{n=0}^{\infty} \mathcal{D}_{2n}$ .

Let  $\mathbf{a} = (a_i)_{i \geq 0}$ ,  $\mathbf{b} = (b_i)_{i \geq 1}$  and  $\mathbf{c} = (c_i)_{i \geq 0}$  be indeterminates; we will work in the ring  $\mathbb{Z}[[\mathbf{a}, \mathbf{b}, \mathbf{c}]]$  of formal power series in these indeterminates. To

each Motzkin path  $\omega$  we assign a weight  $W(\omega) \in \mathbb{Z}[\mathbf{a}, \mathbf{b}, \mathbf{c}]$  that is the product of the weights for the individual steps, where a rise starting at height  $i$  gets weight  $a_i$ , a fall starting at height  $i$  gets weight  $b_i$ , and a level step at height  $i$  gets weight  $c_i$ . Flajolet [104] showed that the generating function of Motzkin paths can be expressed as a continued fraction:

**Theorem 7.6** (Flajolet’s master theorem). *We have*

$$\sum_{\omega \in \mathcal{M}} W(\omega) = \frac{1}{1 - c_0 - \frac{a_0 b_1}{1 - c_1 - \frac{a_1 b_2}{1 - c_2 - \frac{a_2 b_3}{1 - \dots}}}} \quad (7.55)$$

as an identity in  $\mathbb{Z}[[\mathbf{a}, \mathbf{b}, \mathbf{c}]]$ .

In particular, if  $a_{i-1}b_i = \beta_i t^2$  and  $c_i = \gamma_i t$  (note that the parameter  $t$  is conjugate to the length of the Motzkin path), we have

$$\sum_{n=0}^{\infty} t^n \sum_{\omega \in \mathcal{M}_n} W(\omega) = \frac{1}{1 - \gamma_0 t - \frac{\beta_1 t^2}{1 - \gamma_1 t - \frac{\beta_2 t^2}{1 - \dots}}} \quad (7.56)$$

so that the generating function of Motzkin paths with height-dependent weights is given by the J-type continued fraction (7.2). Similarly, if  $a_{i-1}b_i = \alpha_i t$  and  $c_i = 0$  (note that  $t$  is now conjugate to the semi-length of the Dyck path), we have

$$\sum_{n=0}^{\infty} t^n \sum_{\omega \in \mathcal{D}_{2n}} W(\omega) = \frac{1}{1 - \frac{\alpha_1 t}{1 - \frac{\alpha_2 t}{1 - \dots}}} \quad (7.57)$$

so that the generating function of Dyck paths with height-dependent weights is given by the S-type continued fraction (7.1).

Let us now show how to handle Schröder paths within this framework. A *Schröder path* of length  $2n$  ( $n \geq 0$ ) is a path  $\omega = (\omega_0, \dots, \omega_{2n})$  in the right quadrant  $\mathbb{N}_0 \times \mathbb{N}_0$ , starting at  $\omega_0 = (0, 0)$  and ending at  $\omega_{2n} = (2n, 0)$ , whose steps are  $(1, 1)$  [“rise”],  $(1, -1)$  [“fall”] or  $(2, 0)$  [“long level”]. We write  $s_j$  for the step starting at abscissa  $j - 1$ . If the step  $s_j$  is a rise or a fall, we set  $s_j = \omega_j - \omega_{j-1}$  as before. If the step  $s_j$  is a long level step, we set

$s_j = \omega_{j+1} - \omega_{j-1}$  and leave  $\omega_j$  undefined; furthermore, in this case there is no step  $s_{j+1}$ . We write  $h_j$  for the height of the Schröder path at abscissa  $j$  whenever this is defined, i.e.  $\omega_j = (j, h_j)$ . Please note that  $\omega_{2n} = (2n, 0)$  and  $h_{2n} = 0$  are always well-defined, because there cannot be a long level step starting at abscissa  $2n - 1$ . We write  $\mathcal{S}_{2n}$  for the set of Schröder paths of length  $2n$ , and  $\mathcal{S} = \bigcup_{n=0}^{\infty} \mathcal{S}_{2n}$ .

There is an obvious bijection between Schröder paths and Motzkin paths: namely, every long level step is mapped onto a level step. If we apply Flajolet's master theorem with  $a_{i-1}b_i = \alpha_i t$  and  $c_i = \delta_{i+1}t$  to the resulting Motzkin path (note that  $t$  is now conjugate to the semi-length of the underlying Schröder path), we obtain

$$\sum_{n=0}^{\infty} t^n \sum_{\omega \in \mathcal{S}_{2n}} W(\omega) = \frac{1}{1 - \delta_1 t - \frac{\alpha_1 t}{1 - \delta_2 t - \frac{\alpha_2 t}{1 - \dots}}} , \quad (7.58)$$

so that the generating function of Schröder paths with height-dependent weights is given by the T-type continued fraction (7.3). This combinatorial interpretation of T-fractions in terms of Schröder paths was found recently by several authors [110, 164, 133, 188].

## Labelled Schröder paths

Many authors, starting with Flajolet [104], have used bijections from combinatorial objects onto labelled Motzkin or Dyck paths in order to prove J-fraction or S-fraction expansions for the (weighted) ordinary generating functions of those objects. Here we will do the same with labelled Schröder paths in order to prove T-fraction expansions. The definitions are as follows:

Let  $\mathbf{A} = (A_k)_{k \geq 0}$ ,  $\mathbf{B} = (B_k)_{k \geq 1}$  and  $\mathbf{C} = (C_k)_{k \geq 0}$  be sequences of nonnegative integers. An  $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ -labelled Schröder path of length  $2n$  is a pair  $(\omega, \xi)$  where  $\omega = (\omega_0, \dots, \omega_{2n})$  is a Schröder path of length  $2n$ , and  $\xi = (\xi_1, \dots, \xi_{2n})$  is a sequence of integers satisfying

$$1 \leq \xi_i \leq \begin{cases} A(h_{i-1}) & \text{if step } s_i \text{ is a rise (starting at height } h_{i-1}) \\ B(h_{i-1}) & \text{if step } s_i \text{ is a fall (starting at height } h_{i-1}) \\ C(h_{i-1}) & \text{if step } s_i \text{ is a long level step (at height } h_{i-1}) \end{cases} \quad (7.59)$$

[For typographical clarity we have here written  $A(k)$  as a synonym for  $A_k$ , etc.] If step  $s_i$  is undefined (because step  $s_{i-1}$  was a long level step), then  $\xi_i$

is also undefined. We denote by  $\mathcal{S}_{2n}(\mathbf{A}, \mathbf{B}, \mathbf{C})$  the set of  $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ -labelled Schröder paths of length  $2n$ .

Let us stress that the numbers  $A_k$ ,  $B_k$  and  $C_k$  are allowed to take the value 0. Whenever this happens, the path  $\omega$  is forbidden to take a step of the specified kind at the specified height.

We shall also make use of multicoloured Schröder paths. A *k-coloured Schröder path* is simply a Schröder path in which each long level step has been given a “colour” from the set  $\{1, 2, \dots, k\}$ . In other words, we distinguish  $k$  different types of long level steps. An  $(\mathbf{A}, \mathbf{B}, \mathbf{C}^{(1)}, \dots, \mathbf{C}^{(k)})$ -labelled *k-coloured Schröder path of length  $2n$*  is then defined in the obvious way, where we use the sequence  $\mathbf{C}^{(j)}$  to bound the label  $\xi_i$  when step  $i$  is a long level step of type  $j$ . We denote by  $\mathcal{S}_{2n}(\mathbf{A}, \mathbf{B}, \mathbf{C}^{(1)}, \dots, \mathbf{C}^{(k)})$  the set of  $(\mathbf{A}, \mathbf{B}, \mathbf{C}^{(1)}, \dots, \mathbf{C}^{(k)})$ -labelled *k-coloured Schröder paths of length  $2n$* .

## Proof of Theorem 7.3

We will prove Theorem 7.3 by constructing a bijection from the set  $\mathcal{M}_{2n}^*$  of super-augmented perfect matchings of  $[2n]$  onto the set of  $(\mathbf{A}, \mathbf{B}, \mathbf{C}^{(1)}, \mathbf{C}^{(2)})$ -labelled 2-coloured Schröder paths of length  $2n$ , where

$$A_k = 1 \quad \text{for } k \geq 0 \quad (7.60a)$$

$$B_k = k \quad \text{for } k \geq 1 \quad (7.60b)$$

$$C_k^{(1)} = k \quad \text{for } k \geq 0 \quad (7.60c)$$

$$C_k^{(2)} = k + 1 \quad \text{for } k \geq 0 \quad (7.60d)$$

When restricted to ordinary perfect matchings of  $[2n]$  (i.e. super-augmented perfect matchings with no wiggly or dashed lines), our bijection maps onto labelled Dyck paths and coincides with the bijection used by Flajolet [104] and Kasraoui–Zeng [134].<sup>11</sup>

We will begin by explaining how the Schröder path  $\omega$  is defined; then we will explain how the labels  $\xi$  are defined; next we will prove that the mapping is indeed a bijection; next we will translate the various statistics from  $\mathcal{M}_{2n}^*$  to our labelled Schröder paths; and finally we will sum over labels  $\xi$  to obtain the weight  $W(\omega)$  associated to a Schröder path  $\omega$ , which upon applying (7.58) will yield Theorem 7.3.

---

<sup>11</sup> More precisely, Flajolet [104] and Kasraoui–Zeng [134] defined bijections of set partitions of  $[n]$  onto labelled Motzkin paths of length  $n$ . When restricted to perfect matchings of  $[2n]$  (i.e. set partitions of  $[2n]$  in which every block has cardinality 2), their bijections map onto labelled Dyck paths of length  $2n$ . The Flajolet and Kasraoui–Zeng bijections for set partitions are slightly different, but they coincide when restricted to perfect matchings.



**Step 1: Definition of the Schröder path.** Given a super-augmented perfect matching  $\tau \in \mathcal{M}_{2n}^*$ , we define a path  $\omega = (\omega_0, \dots, \omega_{2n})$  starting at  $\omega_0 = (0, 0)$  and ending at  $\omega_{2n} = (2n, 0)$ , with steps  $s_1, \dots, s_{2n}$  ending at locations  $\omega_i = (i, h_i)$ , as follows:

- If  $i$  is a pure opener, then  $s_i$  is a rise, so that  $h_i = h_{i-1} + 1$ .
- If  $i$  is a pure closer, then  $s_i$  is a fall, so that  $h_i = h_{i-1} - 1$ .
- If  $(i, i+1)$  is a wiggly pair, then  $s_i$  is a long level step of type 1 (and  $s_{i+1}$  is undefined). In this case the height  $h_i$  is undefined, but we have  $h_{i+1} = h_{i-1}$ .
- If  $(i, i+1)$  is a dashed pair, then  $s_i$  is a long level step of type 2 (and  $s_{i+1}$  is undefined). In this case the height  $h_i$  is undefined, but we have  $h_{i+1} = h_{i-1}$ .

(See Figure 7.2 for an example.) The interpretation of the heights  $h_i$  is almost immediate from this definition:

**Lemma 7.7.** *For  $i \in \{0, \dots, 2n\}$ , whenever the height  $h_i$  is defined it equals the number of arches that are “started but unfinished” before stage  $i+1$ , i.e.*

$$h_i = \#\{j \leq i < k : j \sim_{\pi(\tau)} k\}. \quad (7.61)$$

In particular, it follows that  $\omega$  is indeed a Schröder path, i.e. all the heights  $h_i$  are nonnegative (when they are defined) and  $h_{2n} = 0$ .

**Step 2: Definition of the labels  $\xi_i$ .**

- If  $i$  is a pure opener, we set  $\xi_i = 1$  as required by (7.60).
- If  $i$  is a pure closer, we look at the  $h_{i-1}$  arches that are “started but unfinished” after stage  $i-1$  (note that we must have  $h_{i-1} \geq 1$ ); let the openers of these arches be  $x_1 < x_2 < \dots < x_{h_{i-1}}$ . Then the vertex  $i$  is paired with precisely one of these openers; if it is  $x_j$ , we set  $\xi_i = j$ . Obviously  $1 \leq \xi_i \leq h_{i-1}$  as required by (7.60).
- If  $(i, i+1)$  is a wiggly pair, then we look at the  $h_{i-1}$  arches that are “started but unfinished” after stage  $i-1$ , and we define  $\xi_i$  exactly as in the preceding case. Again  $1 \leq \xi_i \leq h_{i-1}$ .
- If  $(i, i+1)$  is a dashed pair, then there are two possibilities [see Figure 7.1(b)]: the opener  $i$  and the closer  $i+1$  could belong to different arches (which necessarily cross), or they could belong to the same arch.

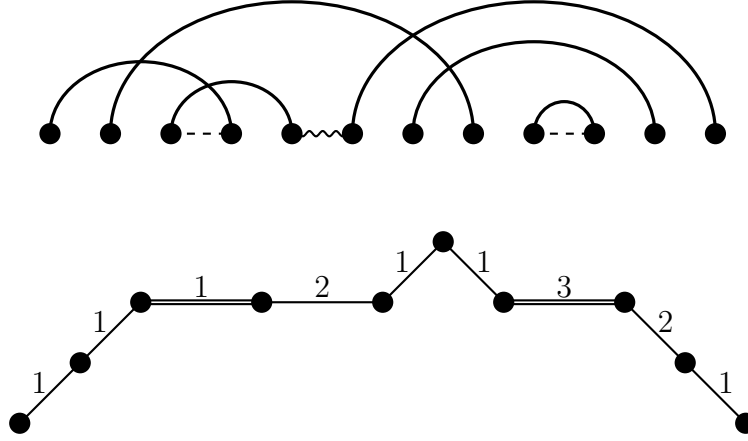


Figure 7.2: A super-augmented perfect matching of  $[2n]$  with  $n = 6$ , together with corresponding labelled Schröder path (the label  $\xi_i$  is written above the step  $s_i$ ). The long level steps of type 2 are shown by double lines.

In the former case we look at the  $h_{i-1}$  arches that are “started but unfinished” after stage  $i - 1$ , and we define  $\xi_i$  as before (except that it is now vertex  $i + 1$  rather than  $i$  that is paired with one of these openers). In the latter case we set  $\xi_i = h_{i-1} + 1$ . Obviously  $1 \leq \xi_i \leq h_{i-1} + 1$  as required by (7.60).

See again Figure 7.2.

**Step 3: Proof of bijection.** It is easy to describe the inverse map from labelled Schröder paths  $(\omega, \xi)$  to super-augmented perfect matchings. Successively for  $i = 1, \dots, n$ , we use the 2-coloured Schröder path  $\omega$  to read off the type associated to step  $s_i$  (opener with no wiggly or dashed line, etc.). And then, if step  $s_i$  corresponds to anything other than an opener with no wiggly or dashed line attached, we use the label  $\xi_i$  to decide to which opener the vertex  $i$  (or  $i + 1$ ) should be attached.

**Step 4: Translation of the statistics.**

**Lemma 7.8.**

(a) If  $i$  is a pure opener, then

$$\text{qne}(i, \pi) = h_{i-1} . \quad (7.62)$$

(b) If  $i$  is a pure closer, then

$$\text{cr}(i, \pi) = h_{i-1} - \xi_i \quad (7.63a)$$

$$\text{ne}(i, \pi) = \xi_i - 1 \quad (7.63b)$$

(c) If  $(i, i+1)$  is a wiggly pair, then also (7.63) holds.

(d) If  $(i, i+1)$  is a dashed pair, then

$$\text{cr}(i+1, \pi) = h_{i-1} + 1 - \xi_i \quad (7.64a)$$

$$\text{ne}(i+1, \pi) = \xi_i - 1. \quad (7.64b)$$

*Proof.* (a) Each of the  $h_{i-1}$  arches that are “started but unfinished” after stage  $i-1$  will either cross or nest the arch that starts at  $i$ ; so this is an immediate consequence of the definition (7.33).

(b,c) Look at the  $h_{i-1}$  arches that are “started but unfinished” after stage  $i-1$ , and let the openers of these arches be  $x_1 < x_2 < \dots < x_{h_{i-1}}$ ; by definition the vertex  $i$  is paired with  $x_{\xi_i}$ . Then each arch starting at a point  $x_j$  with  $j < \xi_i$  must nest with (and lie above) the arch from  $x_{\xi_i}$  to  $i$ , while each arch starting at a point  $x_j$  with  $j > \xi_i$  must cross the arch from  $x_{\xi_i}$  to  $i$ .

(d) Let  $x_1 < x_2 < \dots < x_{h_{i-1}}$  be as before; and let  $x_{h_{i-1}+1} = i$ . By definition the vertex  $i+1$  is paired with  $x_{\xi_i}$ . Then the counting of nestings and crossings is exactly as in (b,c), but with  $h_{i-1}$  replaced by  $h_{i-1} + 1$ .  $\square$

**Step 5: Computation of the weights (7.40).** Using the bijection, we transfer the weights (7.38) from the super-augmented perfect matching to  $(\omega, \xi)$  and then sum over  $\xi$  to obtain the weight  $W(\omega)$ . This weight is factorized over the individual steps  $s_i$ , as follows:

- If  $s_i$  is a rise starting at height  $h_{i-1} = k$  (so that  $i$  is a pure opener), then from (7.62) the weight is

$$a_k = \mathbf{a}_k. \quad (7.65)$$

- If  $s_i$  is a fall starting at height  $h_{i-1} = k$  (so that  $i$  is a pure closer), then from (7.63) the weight is

$$b_k = \sum_{\xi_i=1}^k \mathbf{b}_{k-\xi_i, \xi_i-1} = \mathbf{b}_{k-1}^* \quad (7.66)$$

where  $\mathbf{b}_{k-1}^*$  was defined in (7.41).

- If  $s_i$  is a long level step of type 1 at height  $h_{i-1} = k$  (so that  $(i, i+1)$  is a wiggly pair and  $k \geq 1$ ), then from (7.63) the weight is

$$c_k^{(1)} = \sum_{\xi_i=1}^k f_{k-\xi_i, \xi_i-1} = f_{k-1}^* . \quad (7.67)$$

- If  $s_i$  is a long level step of type 2 at height  $h_{i-1} = k$  (so that  $(i, i+1)$  is a dashed pair), then from (7.64) the weight is

$$c_k^{(2)} = \sum_{\xi_i=1}^{k+1} g_{k+1-\xi_i, \xi_i-1} = g_k^* . \quad (7.68)$$

Setting  $\alpha_i = a_{i-1}b_i$  and  $\delta_i = c_{i-1}^{(1)} + c_{i-1}^{(2)}$  as instructed in (7.58), we obtain the weights (7.40). This completes the proof of Theorem 7.3.

## Bijection between phylogenetic trees and augmented perfect matchings

We will now describe a bijection between augmented perfect matchings of size  $2n$  containing  $k$  wiggly lines and phylogenetic trees with  $n+1$  leaves and  $n-k$  internal vertices. This bijection is illustrated in Figure 7.3.

Given an augmented perfect matching  $P$  of  $[2n]$ , we construct an arch system  $P'$  on  $[2n+1]$  as follows: for each arch  $(i, j)$  in  $P$ , we draw the arch  $(i, j+1)$  in  $P'$  as well as a horizontal line  $(i, i+1)$ . If  $j$  and  $j+1$  are joined by a wiggly line in  $P$ , we make the arch  $(i, j+1)$  wiggly in  $P'$ . Finally, we label the vertices in  $P'$  which are not openers using the numbers  $1, \dots, n+1$ , in order from left to right. Now each vertex  $i \geq 2$  in  $P'$  is joined to exactly one vertex on its left - if  $(i-1)$  is an opener in  $P$ , then  $i$  is joined to  $i-1$  by a horizontal line, whereas if  $i-1$  is a closer in  $P$  then  $i$  is the closer of an arch in  $P'$ . Hence the graph formed by  $P'$  is a tree. Moreover, each opener in  $P'$  is joined to exactly two vertices on its right - one by an arch and one by a horizontal edge. Each labelled vertex in  $P'$  is not joined to any vertices on its right. Hence we can define a planar rooted binary tree  $T_2$  which is isomorphic as a labelled graph to  $P'$ , such that horizontal edges in  $P'$  correspond to left edges in  $T_2$ , while arches in  $P'$  correspond to right edges of  $T_2$  (which are wiggly whenever the corresponding arch is wiggly). Since wiggly edges in the perfect matching join closers to openers, the child of a wiggly edge in  $T_2$  must not be a leaf. Finally, to construct the phylogenetic tree  $T$ , we simply contract all wiggly edges of  $T_2$ .

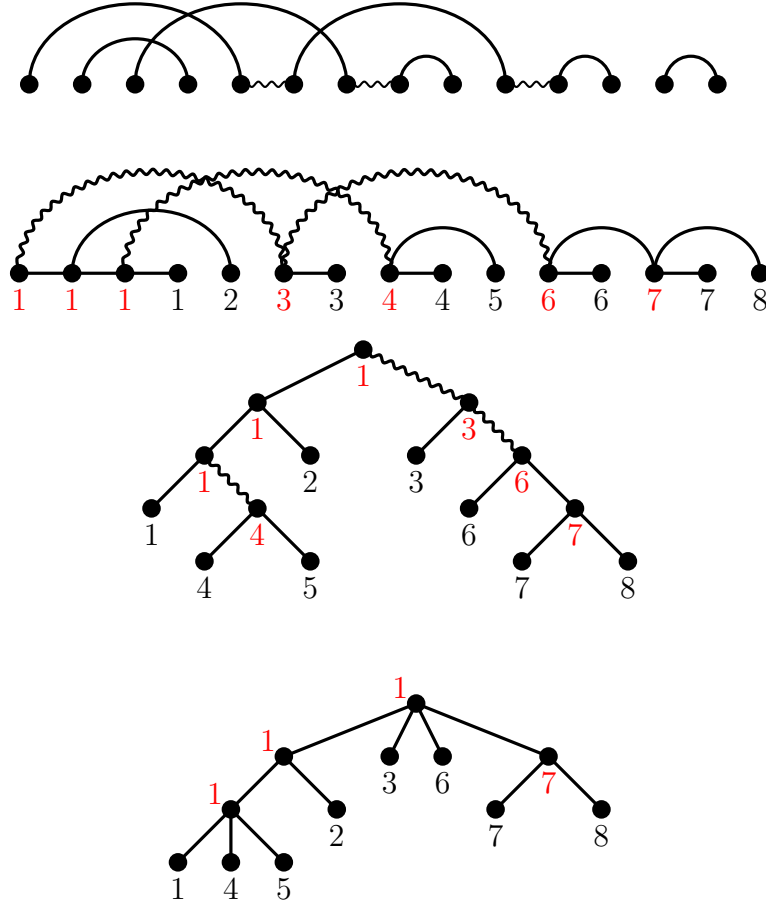


Figure 7.3: A super-augmented perfect matching shown above the corresponding phylogenetic tree, along with the intermediate steps in the bijection. In the forward direction of the bijection, only the black parts are used. The extra labels used in the reverse bijection are shown in red.

To show that this is a bijection, we will describe the reverse transformation. Given a phylogenetic tree  $T$ , we start by recursively labelling each vertex with the minimum label amongst its children. We then order the children of each vertex from left to right in increasing order of their label. It is easy to see that the tree  $T$  constructed from an augmented perfect matching will always be drawn in this way. Now we will describe how to construct the tree  $T_2$  from  $T$ . For each vertex  $v$  with degree at least 3, let  $c_1, c_2, \dots, c_k$  be the children of  $v$  in order from left to right (that is, in increasing order of labels). Then we split  $v$  into a sequence of vertices  $v_1, \dots, v_{k-1}$  of out degree 2 so that the left child of  $v_i$  is  $c_i$  and the right child of  $v_i$  is  $v_{i+1}$  if  $i < k - 1$ , while the right child of  $v_{k-1}$  is  $c_k$ . Moreover, each edge joining  $v_i$  to its right

child  $v_{i+1}$  is wiggly. We then label each new vertex with the same label as its left child. To construct  $P'$  from  $T_2$ , we just have to order the vertices from left to right. If  $u$  and  $v$  are vertices of  $T_2$  such that  $u$  has a lower label than  $v$ , then we say  $u < v$ , and if  $v$  is the left child of  $u$  (so they have the same label) then we say  $u < v$ . We construct  $P'$  by placing the vertices in increasing order according to  $<$ . Finally  $P$  is constructed from  $P'$  by creating an arch  $(i, j)$  for each arch  $(i, j + 1)$  in  $P'$ , and a wiggly line  $(j, j + 1)$  for each wiggly arch  $(i, j + 1)$  in  $P'$ .

Since the transformations between  $P$  and  $T$  are inverses, each transformation is a bijection.

## Recurrence for polynomials defined by the general linear T-fraction

Consider the T-fraction with coefficients  $\alpha_i = x + (i - 1)u$  and  $\delta_i = z + (i - 1)w$ , and let  $P_n(x, u, z, w)$  be the polynomials that it generates:

$$f(t; x, u, z, w) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} P_n(x, u, z, w) t^n \quad (7.69a)$$

$$\stackrel{\text{def}}{=} \frac{1}{1 - zt - \frac{xt}{1 - (z + w)t - \frac{(x + u)t}{1 - (z + 2w)t - \frac{(x + 2u)t}{1 - \dots}}}} \quad (7.69b)$$

**Proposition 7.9.** *The ordinary generating function  $f(t; x, u, z, w)$  satisfies the nonlinear partial differential equation*

$$f = 1 + (z + u)tf + ut^2f_t + (u + w)t(uf_u + xf_x) + (x - u)tf^2. \quad (7.70)$$

*Equivalently, the polynomials  $P_n(x, u, z, w)$  satisfy the non-linear recurrence*

$$P_n = (z + nu)P_{n-1} + (u + w)\left(u\frac{\partial P_{n-1}}{\partial u} + x\frac{\partial P_{n-1}}{\partial x}\right) + (x - u)\sum_{j=0}^{n-1} P_j P_{n-1-j}. \quad (7.71)$$

In particular, when we restrict to  $u = x$  the polynomials  $P_n(x, x, z, w)$  satisfy the linear recurrence

$$P_n = (z + nx)P_{n-1} + x(x + w)\frac{\partial P_{n-1}}{\partial x}. \quad (7.72)$$

When we restrict to  $w = -u$  the generating function  $f(t; x, u, z, -u)$  satisfies the Riccati type equation

$$f(t) = 1 + (z + u)tf(t) + ut^2f'(t) + (x - u)tf(t)^2.$$

We will now prove Proposition 7.9 by choosing particular values of the weights in Theorem 7.3 so that the generating function is equal to  $f(t; x, u, z, w)$ , then proceed to prove (7.70) combinatorially.

*Proof.* In Theorem 7.3 we set  $f_{l,l'} = 0$ ,  $a_l = 1$ ,  $g_{0,l'} = z$ ,  $b_{0,l'} = x$ . For  $l > 0$ , we set  $g_{l,l'} = w$  and  $b_{l,l'} = u$ . Then the generating function

$$\sum_{n=0}^{\infty} M_n t^n$$

is equal to the T-fraction with coefficients  $\alpha_i = x + (i - 1)u$  and  $\delta_i = z + (i - 1)w$ , which is exactly the form of  $f(t; x, u, z, w)$ . Hence

$$f(t; x, u, z, w) = \sum_{n=0}^{\infty} M_n t^n,$$

and  $M_n = P_n(x, u, z, w)$  for each  $n$ . Moreover, by the definition of the polynomials  $M_n$ , the series  $f(t; x, u, z, w)$  is the generating function of augmented perfect matchings with no wiggly lines, where  $t$  counts arches,  $z$  counts dashed lines where the two ends are part of the same arch,  $w$  counts other dashed lines,  $x$  counts pure closers with crossing number 0 and  $u$  counts pure closers with crossing number greater than 0. We will now prove combinatorially that  $f(t; x, u, z, w)$  satisfies (7.70).

The contribution to  $f$  from the case with no arches is clearly 1. Otherwise, let the arch whose opener is the initial vertex be called  $\alpha$ , and we call the closer of this arch  $c_\alpha$ . Let  $A'$  be the perfect matching which remains when  $\alpha$ , its incident vertices and their incident dashed lines are removed, and let  $t^{a+b+c+d}x^a u^b z^c w^d$  be the contribution of  $A'$  to  $F$ . We will proceed by considering five cases for the type of  $\alpha$  illustrated in Figure 7.4. Case (i) is when the initial vertex is incident on a dashed line. Then the second point is a closer, so it must close  $\alpha$ . Then  $A'$  can be any dash-augmented arch system, so the contribution from this case is  $ztF$ .

In the remaining four cases, the initial vertex is a pure opener. Case (ii) is when  $c_\alpha$  is incident on a dashed line. Then  $\alpha$  contributes the weight  $wt$ . The closer  $c_\alpha$  can be placed immediately after any of the  $a + b$  pure openers in  $A'$ . Hence the contribution to  $F$  from arch systems corresponding to the smaller arch system  $A'$  is  $(a + b)t^{a+b+c+d+1}x^a u^b z^c w^{d+1}$ . Summing this over

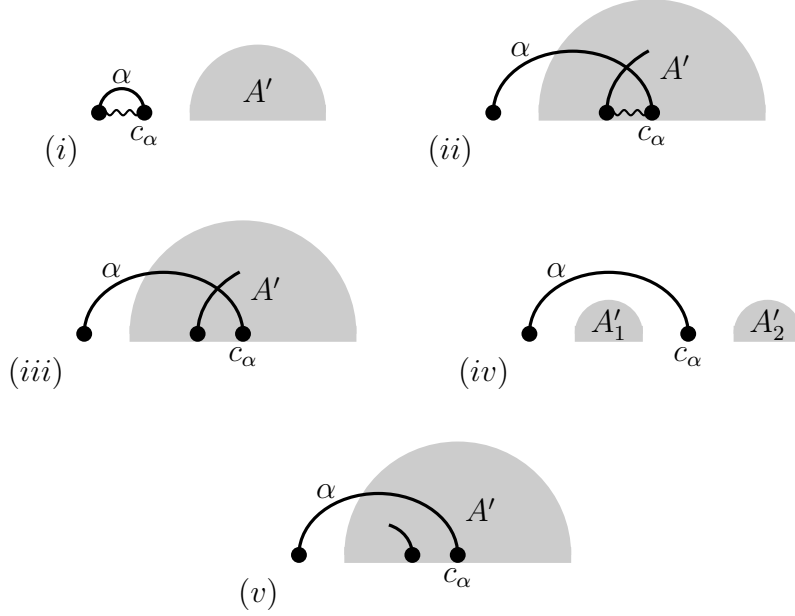


Figure 7.4: The five cases in the proof of Proposition 7.9.

all possible arch systems  $A'$  yields the contribution  $wt(uF_u + xF_x)$  from this case.

In the remaining three cases,  $c_\alpha$  is a pure closer. Case (iii) is when  $c_\alpha$  immediately follows the opener of a different arch. This is identical to Case 2 except that we do not attach a dashed line to  $c_\alpha$ , so  $\alpha$  contributes the weight  $ut$  instead of  $wt$ . Hence the contribution from this case is  $ut(uF_u + xF_x)$ .

Case (iv) is when  $\alpha$  has crossing number 0. Then the arch system  $A'$  separates into two sections  $A'_1$  and  $A'_2$ , shown in Figure 7.4, where  $A'_1$  is the section contained under  $\alpha$  and  $A'_2$  is the section which follows  $c_\alpha$ . The arch  $\alpha$  contributes the weight  $xt$ , so the contribution from this case is  $xtF^2$ .

Case (v) is when  $\alpha$  has crossing number greater than 0 and  $c_\alpha$  immediately follows another closer. If we remove the condition that  $\alpha$  has crossing number greater than 0, then  $c_\alpha$  can follow any closer in  $A'$ , so there are  $a + b + c + d$  possible positions for  $c_\alpha$ . Ignoring the weight contributed by  $\alpha$ , these are counted by  $tF_t$ . Now we subtract  $F^2$  from this to remove the cases where  $\alpha$  has crossing number 0, however we add back  $F$  to account for the case in which  $c_\alpha$  immediately follows the initial vertex. multiplying by the weight  $ut$  of  $\alpha$  yields the contribution  $ut(tF_t - F^2 + F)$  for this case.

Adding the contributions from all 5 cases yields the desired result. This concludes the proof of Proposition 7.9.  $\square$



## Chapter 8

# Off-critical parafermions and the winding angle distribution of the $O(n)$ model

### Introduction

The  $n$ -vector model, introduced by Stanley in 1968 [192] is described by the Hamiltonian

$$\mathcal{H}(d, n) = -J \sum_{\langle i, j \rangle} \mathbf{s}_i \cdot \mathbf{s}_j,$$

where  $d$  denotes the dimensionality of the lattice, and  $\mathbf{s}_i$  is an  $n$ -dimensional unit vector. When  $n = 1$  this Hamiltonian describes the Ising model, when  $n = 2$  it describes the classical XY model, and in the limit  $n \rightarrow 0$ , one recovers the self-avoiding walk (SAW) model, as first pointed out by de Gennes [114]. Of particular relevance here is the fact that the  $n$ -vector model on the honeycomb lattice has been shown [76] to be equivalent to a loop model with a weight  $n$  attached to closed loops.

In 1982 Nienhuis [162] showed that, for  $n \in [-2, 2]$ , the model on the honeycomb lattice could be mapped onto a solid-on-solid model, from which he was able to derive the critical points and critical exponents, subject to some plausible assumptions. These results agreed with the known exponents and critical point for the Ising model, and predicted exact values for those models corresponding to other values of the spin dimensionality  $n$ . In particular, for  $n = 0$  the critical point for the honeycomb lattice SAW model was predicted to be  $x_c = 1/\sqrt{2 + \sqrt{2}}$ , a result finally proved 28 years later by Duminil-Copin and Smirnov [79] using an identity that follows from discrete analogues of Cauchy-Riemann conditions for a parafermionic observ-

able. This identity holds only at the critical point  $x_c$ , so cannot say anything about exponents, which require knowledge of the behaviour as  $x \rightarrow x_c$ .

Smirnov [185] has also derived such an identity for the general honeycomb  $O(n)$  model with  $n \in [-2, 2]$ . This identity provides an alternative way of predicting the value of the critical point  $x_c(n) = 1/\sqrt{2 + \sqrt{2 - n}}$  as conjectured by Nienhuis for values of  $n$  other than  $n = 0$ .

This chapter contains two new results. We first present an off-critical deformation of the discrete Cauchy-Riemann conditions, which allows us to consider critical exponents near criticality. Indeed, this deformation gives rise to an identity between bulk and boundary generating functions, and we utilise this identity to determine the asymptotic form of the winding angle distribution function for SAWs on the half-plane and in a wedge in terms of boundary critical exponents. It is important to note that the only assumptions we make are the existence of the critical exponents and the value of the critical point. We will not rely on Coulomb gas techniques or conformal invariance. We find perfect agreement with the conjectured winding angle distribution function on the cylinder predicted by Duplantier and Saleur [81] in terms of bulk critical exponents. These predictions of Duplantier and Saleur were reliant on the validity of Coulomb-gas techniques and conformal invariance.

## Off-critical identity for the honeycomb $O(n)$ model

We define a *domain*  $\Omega$  to be a simply connected collection of mid-edges on a half-plane of the honeycomb lattice. The set of vertices adjacent to the mid-edges of  $\Omega$  is denoted  $V(\Omega)$ . Those mid-edges of  $\Omega$  which are adjacent to only one vertex in  $V(\Omega)$  form the boundary  $\partial\Omega$ .

Let  $\gamma$  be a configuration on a domain  $\Omega$  comprising a single self-avoiding walk and a number (possibly zero) of closed loops. We denote by  $\ell(\gamma)$  the number of vertices occupied by  $\gamma$  and  $c(\gamma)$  the number of closed loops. Let furthermore  $W(\gamma)$  be the winding angle of the self-avoiding walk component. Define the following observable:

**Definition 8.1** (Preholomorphic observable).

- For  $a \in \partial\Omega, z \in \Omega$ , set

$$F(\Omega, z; x, n, \sigma) := F(z) = \sum_{\substack{\gamma: a \rightarrow z \\ a, z \in \Omega}} e^{-i\sigma W(\gamma)} x^{\ell(\gamma)} n^{c(\gamma)}, \quad (8.1)$$

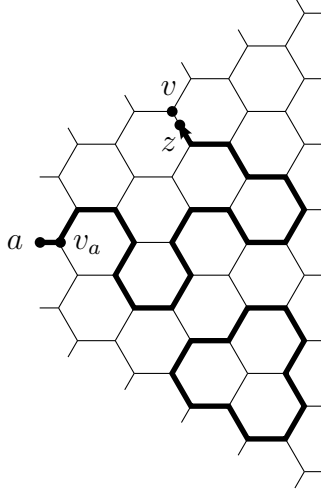


Figure 8.1: A configuration  $\gamma$  on a finite domain. Points  $a$  and  $z$  are mid-edges and  $v_a$  and  $v$  are corresponding vertices. The contribution of  $\gamma$  to  $F(z)$  is  $e^{-2i\sigma\pi/3}x^{30}n$ .

where the sum is over all configurations  $\gamma$  for which the SAW component runs from the mid-edge  $a$  to a mid-edge  $z$  (we say that  $\gamma$  ends at  $z$ ).

- Let  $F(p; v)$  only include configurations where there is a walk terminating at the mid-edge  $p$  and heading in the direction of the vertex  $v$ . For  $v_a, v \in V(\Omega)$ , set

$$\begin{aligned} F(V(\Omega), v; x, n, \sigma) &:= F(v) = (p - v)F(p; v) + (q - v)F(q; v) + (r - v)F(r; v) \\ &= - \sum_{\substack{\gamma: v_a \rightarrow v \\ v_a, v \in V(\Omega)}} e^{i\tilde{\sigma}W(\gamma)} x^{\ell(\gamma)} n^{c(\gamma)}, \end{aligned} \quad (8.2)$$

where  $\tilde{\sigma} = 1 - \sigma$ . The sum in the second line of (8.2) is over all configurations  $\gamma$  for which the SAW component runs from vertex  $v_a$  to a vertex  $v$ .

See Figure 8.1 for an example.

Smirnov [185] proves the following.

**Lemma 8.2** (Smirnov). *For  $n \in [-2, 2]$ , set  $n = 2 \cos \phi$  with  $\phi \in [0, \pi]$ . Then for*

$$\sigma = \frac{\pi - 3\phi}{4\pi}, \quad x_c^{-1} = 2 \cos \left( \frac{\pi + \phi}{4} \right) = \sqrt{2 - \sqrt{2 - n}}, \quad \text{or} \quad (8.3)$$

$$\sigma = \frac{\pi + 3\phi}{4\pi}, \quad x_c^{-1} = 2 \cos \left( \frac{\pi - \phi}{4} \right) = \sqrt{2 + \sqrt{2 - n}}, \quad (8.4)$$

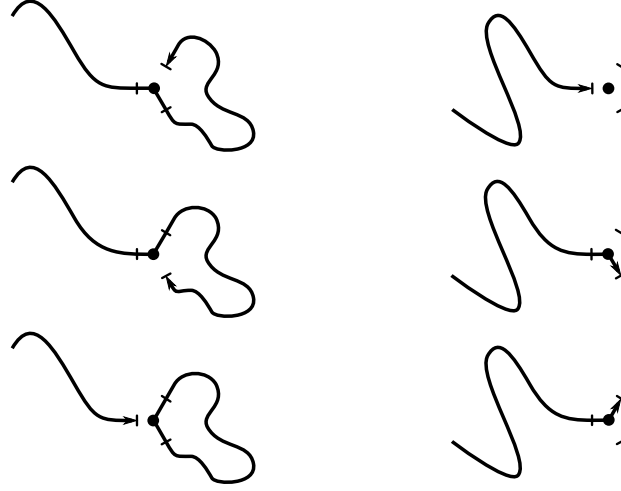


Figure 8.2: The two ways of grouping the configurations which end at mid-edges  $p, q, r$  adjacent to vertex  $v$ . On the left, configurations which visit all three mid-edges; on the right, configurations which visit one or two of the mid-edges.

the observable  $F$  satisfies the following relation for every vertex  $v \in V(\Omega)$ :

$$(p - v)F(p) + (q - v)F(q) + (r - v)F(r) = 0, \quad (8.5)$$

where  $p, q, r$  are the mid-edges of the three edges adjacent to  $v$ .

The first equation in Lemma 8.2 corresponds to the larger of the two critical values of the step weight  $x$  and thus describes the so-called dense regime as configurations with many loops are favoured. The second equation corresponds to the line of critical points separating the dense and dilute phases [162]. The proof of this lemma is well known, but we include it here for later reference.

*Proof.* The lemma follows from considering contributions around the vertex  $v$  (see Figure 8.2). There are two possible cases. In the following, we define  $\lambda = e^{-i\sigma\pi/3}$  (the weight accrued by a walk for each left turn) and  $j = e^{2i\pi/3}$  (the value of  $p - v$  when mid-edge  $p$  is to the north-west of its adjacent vertex  $v$ ).

- (a) In the first case, all mid-edges  $p, q, r$  are visited by a configuration and hence two of the three edges incident on  $v$  must be occupied. There are three ways for this to occur: two with the self-avoiding walk visiting all three sites, and one with a closed loop running through  $v$ . If we fix the rest of the configuration not incident on  $v$ , the three contributions

add up to zero if

$$\bar{j}\lambda^4 + j\bar{\lambda}^4 + n = 0. \quad (8.6)$$

This equation determines the possible values of the parameter  $\sigma$ . There are two solutions:

$$\sigma = \frac{\pi - 3\theta}{4\pi} \quad \text{for } \bar{j}\lambda^4 = -e^{i\phi}, \quad (8.7)$$

$$\sigma = \frac{\pi + 3\theta}{4\pi} \quad \text{for } \bar{j}\lambda^4 = -e^{-i\phi}. \quad (8.8)$$

- (b) In the second case only one or two mid-edges are occupied in the configuration, and the condition that all contributions add up to zero becomes

$$1 + x_c j \bar{\lambda} + x_c \bar{j} \lambda = 0, \quad (8.9)$$

which leads to

$$x_c^{-1} = 2 \cos \left( \frac{\pi}{3} (\sigma - 1) \right). \quad (8.10)$$

The two possible values of  $\sigma$  give rise to the corresponding two values for  $x_c$ .  $\square$

## Off-critical deformation

First we evaluate the discrete divergence of the second set of configurations in Figure 8.1 for general  $x$ , below the critical value. This gives

**Lemma 8.3** (Massive preholomorphicity condition). *For a given vertex  $v$  with mid-edges  $p$ ,  $q$  and  $r$ , and  $x$  below the critical value  $x_c$ , the parafermionic observable  $F(z)$  satisfies*

$$(p - v)F(p) + (q - v)F(q) + (r - v)F(r) = \left(1 - \frac{x}{x_c}\right)F(v) \quad (8.11)$$

where  $F(z)$  when  $z$  is a mid-edge, and  $F(v)$  when  $v$  is a vertex are defined in Definition 8.1.

We use the term massively preholomorphic as (8.11) is of a similar form to that described in [152] and [185].

*Proof.* Similar to Lemma 8.2 the proof splits into two parts. The first part, concerning cancellations of contributions coming from walks depicted in the left hand side of Figure 8.3, is completely analogous. The second part however changes.

Consider a vertex  $v$  with mid-edges labelled  $p$ ,  $q$  and  $r$  in a counter-clockwise fashion. There are three disjoint sets of configurations, depending on which of the three mid-edges  $p$ ,  $q$  or  $r$  the walk enters from. These are shown in Figure 8.3. Denote by  $F(p; v)$  the contributions to  $F(p)$  that only include configurations where there is a walk terminating at the mid-edge  $p$  heading in the direction of the vertex  $v$ . For walks entering the vertex from  $p$ , the left-hand side of equation (8.11) becomes

$$\begin{aligned} & (p - v)F(p; v) + e^{2\pi i/3}(p - v)x e^{\pi\sigma i/3}F(p; v) + e^{-2\pi i/3}(p - v)x e^{-\pi\sigma i/3}F(p; v) \\ = & (p - v)F(p; v)(1 - x e^{\pi i/3(\sigma-1)} - x e^{-\pi i/3(\sigma-1)}) \\ = & (p - v)F(p; v)(1 - \frac{x}{x_c}) \end{aligned}$$

where in the first line we have used that

$$q - v = e^{2\pi i/3}(p - v), \quad r - v = e^{-2\pi i/3}(p - v),$$

For walks entering from mid-edges  $q$  and  $r$  similar calculations give contributions

$$(q - v)F(q; v)(1 - \frac{x}{x_c}) \text{ and } (r - v)F(r; v)(1 - \frac{x}{x_c}).$$

Adding the three contributions together gives the right-hand side of equation (8.11).  $\square$

Using the above lemma we can now derive the following relationship between generating functions.

**Lemma 8.4** (Off-critical generating function identity).

$$\sum_{\gamma: a \rightarrow \partial\Omega} e^{i\tilde{\sigma}W(\gamma)} x^{|\gamma|} n^{c(\gamma)} + (1 - x/x_c) \sum_{\gamma: a \rightarrow \Omega \setminus \partial\Omega} e^{i\tilde{\sigma}W(\gamma)} x^{|\gamma|} n^{c(\gamma)} = C_\Omega(x) \quad (8.12)$$

where

$$C_\Omega(x) = \sum_{\gamma: a \rightarrow a} x^{|\gamma|} n^{c(\gamma)} \quad (8.13)$$

is the usual generating function of the honeycomb lattice  $O(n)$  model, i.e. closed loops without the SAW component, and  $\tilde{\sigma} = 1 - \sigma$ .

*Proof.* We begin by summing equation (8.11) over all the vertices of the lattice  $\Omega$ . As in the critical case, there is cancellation at points in the bulk

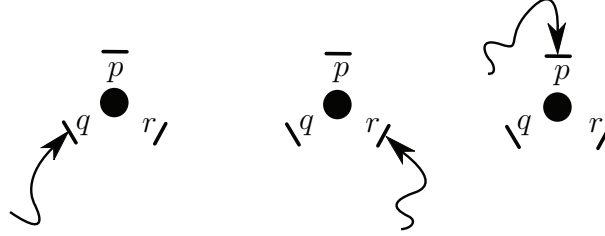


Figure 8.3: The three possible ways for a walk to enter a given vertex via each of the three mid-edges,  $p$ ,  $q$  and  $r$ . The discrete divergence is evaluated for all three cases in order to derive the off-critical, or ‘massive’ preholomorphicity condition.

so that the left-hand side contains only boundary terms and we have the following equation

$$\begin{aligned} & \sum_{\gamma: a \rightarrow \partial\Omega} e^{i\tilde{\sigma}W(\gamma)} x^{|\gamma|} n^{c(\gamma)} - \sum_{\gamma: a \rightarrow a \in \partial\Omega} x^{|\gamma|} n^{c(\gamma)} \\ &= \left(1 - \frac{x}{x_c}\right) \sum_{\gamma: a \rightarrow z} (F(z; v_1)(z - v_1(z)) + F(z; v_2)(z - v_2(z))) \end{aligned} \quad (8.14)$$

Note that there are two terms in the summation on the right-hand side. This is because for a given end point  $z$ , a walk can be heading towards one of two possible vertices which we call  $v_1$  and  $v_2$ , the labelling being unimportant. This is illustrated in Figure 8.4. Using  $\sigma = 1 - \tilde{\sigma}$  and the definition of  $F(z; v)$  the right hand side becomes, up to a factor  $(1 - x/x_c)$ ,

$$e^{i\phi} \left( \sum_{\gamma: a \rightarrow z \rightarrow v_1} x^{|\gamma|} n^{c(\gamma)} e^{i\tilde{\sigma}W(\gamma)} e^{-iW(\gamma)} - \sum_{\gamma: a \rightarrow z \rightarrow v_2} x^{|\gamma|} n^{c(\gamma)} e^{i\tilde{\sigma}W(\gamma)} e^{-iW(\gamma)} \right)$$

where  $e^{i\phi}$  is the unit vector from  $v_1$  to  $z$ , which is the negative of the unit vector from  $v_2$  to  $z$ .

A walk that terminates at  $z$  and moves in the direction of vertex  $v_2$  has winding  $W(\gamma_2) = 2\pi m' + \phi$  while a walk heading in the direction of the vertex  $v_1$  has winding  $W(\gamma_1) = (2m + 1)\pi + \phi$  for some  $m, m' \in \mathbb{Z}$ . In each case the angle  $\phi$  from the unit vector is cancelled by the  $\phi$  appearing in the winding angle term  $e^{-iW(\gamma)}$  and this leaves

$$- \sum_{\gamma: a \rightarrow z} x^{|\gamma|} n^{c(\gamma)} e^{i\tilde{\sigma}W(\gamma)}. \quad (8.15)$$

The left-hand side is a sum of walks to the boundary and a sum over configurations with walks of length zero, which is  $C_\Omega(x)$ . Substituting expression



Figure 8.4: A walk terminating at the mid-edge  $z$ . The mid-edge lies between two vertices  $v_1$  and  $v_2$  and the unit vector from  $v_1$  to  $z$  is given by  $e^{i\phi}$ . The labelling of the vertices is arbitrary.

(8.15) into equation (8.14) and rearranging we obtain

$$\sum_{\gamma: a \rightarrow \partial\Omega} e^{i\tilde{\sigma}W(\gamma)} x^{|\gamma|} n^{c(\gamma)} + (1 - x/x_c) \sum_{\gamma: a \rightarrow \Omega \setminus \partial\Omega} e^{i\tilde{\sigma}W(\gamma)} x^{|\gamma|} n^{c(\gamma)} = C_\Omega(z). \quad (8.16)$$

□

## Winding angle

Let us now restrict to a particular trapezoidal domain  $\Omega = S_{T,L}$  of width  $T$  and left-height  $2L$ , see Figure 8.5. The winding angle distribution function can be calculated directly from the off-critical generating function identity (8.16). First define the following function

$$G_{\theta,\Omega}(x) = \sum_{\substack{\gamma: a \rightarrow \Omega \setminus \partial\Omega \\ W(\gamma) = \theta}} x^{|\gamma|} n^{c(\gamma)}.$$

$G_{\theta,\Omega}(x)$  contains only those contributions to  $G_\Omega(x) = \sum_\theta G_{\theta,\Omega}(x)$  that involve a walk with winding angle  $\theta$ . Also, define

$$H_\Omega(x) = \sum_{\gamma: a \rightarrow \partial\Omega} e^{i\tilde{\sigma}W(\gamma)} x^{|\gamma|} n^{c(\gamma)},$$

which is the generating function of walks that end on a boundary of the domain, and thus have a winding angle associated to that boundary. We remind the reader here that  $\gamma$  describes a walk along with a configuration of loops. Using this notation (8.16) becomes

$$H_\Omega(x) + (1 - x/x_c) \sum_\theta e^{i\tilde{\sigma}\theta} G_{\theta,\Omega}(x) = C_\Omega(x).$$

Now let  $H_\Omega^*(x)$  and  $G_{\theta,\Omega}^*(x)$  be  $H_\Omega(x)/C_\Omega(x)$  and  $G_{\theta,\Omega}(x)/C_\Omega(x)$  respectively. For  $x < x_c$  define  $H^*(x)$  and  $G_\theta^*(x)$  to be  $H_\Omega^*(x)$  and  $G_{\theta,\Omega}^*(x)$  respectively in



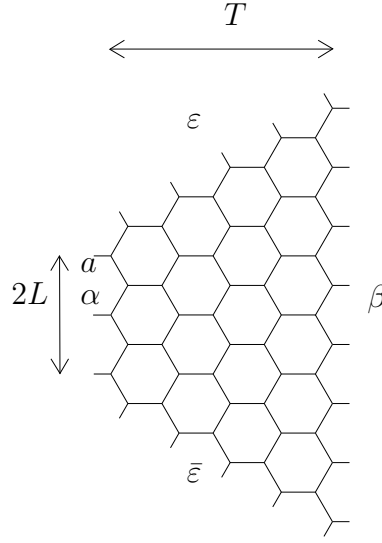


Figure 8.5: Finite patch  $S_{3,1}$  of the hexagonal lattice. The SAW component of a loop configuration starts on the central mid-edge of the left boundary (shown as  $a$ ).

the limit as  $\Omega$  approaches the half plane. That is to say, we have first let the length of the domain  $L$  become infinite, so that the domain becomes a strip, and then let the strip width  $T$  go to infinity. Thus we are discussing walks attached to the surface  $\alpha$ , which is the only surface remaining in the domain. This limit is perfectly well defined. The original domain was trapezoidal, with left-height  $2L$  and width  $T$ . In [18] we showed that the generating function for walks that finish at the top and bottom of the trapezoid vanish as  $L \rightarrow \infty$ . In the resulting strip geometry there are two generating functions describing walks that end on a surface, one corresponding to the surface where the walks originate and the other corresponding to the parallel surface at distance  $T$ . In [79] it is shown that this generating function is bounded above by  $(x/x_c)^T$ . Since  $x < x_c$ , this upper bound vanishes as  $T \rightarrow \infty$ . Therefore in the limit as the strip width becomes infinite we obtain

$$H^*(x) + (1 - x/x_c) \sum_{\theta} e^{i\tilde{\sigma}\theta} G_{\theta}^*(x) = 1. \quad (8.17)$$

## Susceptibilities and critical exponents

The first term  $H^*(x)$  is (up to a multiplicative constant) the normalised generating function of walks that start and end at the  $\alpha$  surface (and have an additional half-step to the left of their starting and ending vertices). This

generating function is usually denoted in the literature as  $\chi_{11}(x)$  [29]. One conventionally writes

$$\chi_{11}(x) \sim d_0(x) + d_1(x)(1 - x/x_c)^{-\gamma_{11}}, \quad x \lesssim x_c,$$

where  $d_0$  and  $d_1$  are analytic near  $x = x_c$ . This assumes the existence of the exponent  $\gamma_{11}$  and that the value of the critical exponent is given by  $x_c$  as defined in (8.4).<sup>1</sup>

Similarly, the generating function of walks that start at the surface and finish anywhere inside the domain, is usually denoted in the literature as  $\chi_1(x)$ . One conventionally writes

$$\chi_1(x) \sim c(x)(1 - x/x_c)^{-\gamma_1}, \quad x \lesssim x_c,$$

with  $c(x)$  analytic. This assumes the existence of the exponent  $\gamma_1$  and the value of the critical point, but no other assumption is made. Note that this assumption for example is not valid at  $n = 2$ .

With the assumption as to the existence of the exponents for the susceptibilities we thus have

$$H^*(x) \propto \chi_{11}(x) \sim 1 + \text{const} \times (1 - x/x_c)^{-\gamma_{11}},$$

and

$$\sum_{\theta} G_{\theta}^*(x) \propto \chi_1(x) \sim \text{const} \times (1 - x/x_c)^{-\gamma_1}, \quad (8.18)$$

Using equation (8.17) we also obtain

$$\sum_{\theta} e^{\tilde{\sigma}i\theta} G_{\theta}^*(x) = (1 - H^*(x))(1 - x/x_c)^{-1} \sim \text{const} \times (1 - x/x_c)^{-\gamma_{11}-1}. \quad (8.19)$$

## Asymptotic winding angle distribution

Let us first recall a standard result for series near their radius of convergence:

**Lemma 8.5.** *Let  $G(x) = \sum_{j \geq 0} g_j x^j$ . If  $G(x) \sim A(1 - x/x_c)^{-\eta}$  for  $x \lesssim x_c$  and some constant  $A$ , then*

$$g_j \sim A x_c^{-j} j^{\eta-1} / \Gamma(\eta).$$

We denote the number of walks of length  $j$  with winding angle exactly  $\theta$  by  $a_{\theta}(j)$ , so that we can write

$$G_{\theta}^*(x) = \sum_{j=0}^{\infty} a_{\theta}(j) x^j.$$

---

<sup>1</sup>For SAWs ( $n = 0$ ) it was proved in [79] that  $x_c$  is indeed the critical point.

Summing over  $\theta$  and using (8.18) and Lemma 8.5, the total number of walks of length  $j$  behaves like

$$\sum_{\theta} a_{\theta}(j) \sim \text{const} \times x_c^{-j} j^{\gamma_1-1}.$$

Similarly, from (8.19) we have

$$\sum_{\theta} e^{i\tilde{\sigma}\theta} a_{\theta}(j) \sim \text{const} \times x_c^{-j} j^{\gamma_{11}}.$$

**Definition 8.6.** The probability density function  $P(\theta, j)$  for the winding angle of walks of length  $j$  is the fraction of walks of length  $j$  with winding angle  $\theta$ :

$$P(\theta, j) = \frac{a_{\theta}(j)}{\sum_{\theta} a_{\theta}(j)}.$$

From the reasoning above, the following result follows immediately.

**Proposition 8.7.** *Let  $\tilde{\sigma} = 1 - \sigma$  where  $\sigma$  is given by (8.8). Then*

$$\sum_{\theta} e^{i\tilde{\sigma}\theta} P(\theta, j) \sim \text{const} \times j^{\gamma_{11}-\gamma_1+1}.$$

That is to say, the probability density function is characterised by an exponent that can be expressed solely in terms of the half-plane exponents  $\gamma_1$  and  $\gamma_{11}$ .

## Winding angle in a wedge

Rather than taking  $L \rightarrow \infty$  as in the previous sections, we now set  $L = 0$ . In this case the trapezoidal domain  $S_T, 0$  reduces to a wedge with wedge angle  $\alpha = \pi/3$ . Using similar arguments as before, we take the limit  $T \rightarrow \infty$ , giving

$$H_{\alpha}^*(x) + (1 - x/x_c) \sum_{\theta} e^{i\tilde{\sigma}\theta} G_{\theta, \alpha}^*(x) = 1. \quad (8.20)$$

where now  $H_{\alpha}^*(x)$  is (up to a multiplicative constant) the generating function of walks that start at  $a$  and end at the  $\varepsilon$  or  $\bar{\varepsilon}$  surface of the wedge with angle  $\alpha = \pi/3$  (see Figure 8.5). This generating function is usually denoted in the literature  $\chi_{21}(x, \alpha)$  [127]. It is known that self-avoiding walks ( $n = 0$ ) in a wedge have the same connective exponent as those in the plane, for arbitrary wedge angle [129]. This is also known to be true in the Ising ( $n = 1$ ) case

[13], and for the  $O(-2)$  model, the Gaussian case. We assume that this holds also for all  $n \in [-2, 2]$ . We thus write

$$\chi_{21}(x, \pi/3) \sim \tilde{d}_0(x) + \tilde{d}_1(x)(1 - x/x_c)^{-\gamma_{21}(\pi/3)},$$

where  $\tilde{d}_0$  and  $\tilde{d}_1$  are analytic near  $x = x_c$ . This assumes the existence of the exponent  $\gamma_{21}(\alpha)$  which in general depends on the wedge angle  $\alpha$ . Similarly, the generating function of walks that start at  $a$  and finish anywhere inside the domain is usually denoted in the literature as  $\chi_2(x, \alpha)$ . Assuming the existence of the relevant critical exponent  $\gamma_2(\alpha)$ , we write

$$\chi_2(x, \pi/3) \sim \tilde{c}(x)(1 - x/x_c)^{-\gamma_2(\pi/3)},$$

with  $\tilde{c}(x)$  analytic near  $x = x_c$ .

Denote the probability density function for the winding angle of walks of length  $j$  in a wedge with angle  $\alpha$  by  $P_\alpha(\theta, j)$ . Using exactly the same reasoning as in the previous section, we find that

$$\sum_{\theta} e^{i\tilde{\sigma}\theta} P_{\alpha=\pi/3}(\theta, j) \sim \text{const} \times j^{\gamma_{21}(\pi/3) - \gamma_2(\pi/3) + 1}. \quad (8.21)$$

In fact, one can approximate every wedge angle by an appropriate stacking of honeycombs. Using the exact same reasoning as before we can therefore generalise to arbitrary wedge angle and find:

$$\sum_{\theta} e^{i\tilde{\sigma}\theta} P_{\alpha}(\theta, j) \sim \text{const} \times j^{\gamma_{21}(\alpha) - \gamma_2(\alpha) + 1}. \quad (8.22)$$

This reduces to the previous result in the special case  $\alpha = \pi$ .

## Exponent inequalities

Using the techniques employed above we can derive rigorous exponent bounds in the following way. Recall

$$H^*(x) \propto \chi_{11}(x) \sim 1 + \text{const} \times (1 - x/x_c)^{-\gamma_{11}},$$

and

$$\sum_{\theta} G_{\theta}^*(x) \propto \chi_1(x) \sim \text{const} \times (1 - x/x_c)^{-\gamma_1}.$$

Since

$$\sum_{\theta} e^{i\tilde{\sigma}\theta} G_{\theta}^*(x) \leq \sum_{\theta} G_{\theta}^*(x),$$

it follows from

$$H^*(x) + (1 - x/x_c) \sum_{\theta} e^{i\tilde{\sigma}\theta} G_{\theta}^*(x) = 1,$$

that

$$\gamma_1 - \gamma_{11} \geq 1. \quad (8.23)$$

The only assumptions here are the existence of the critical exponents and, for  $n \neq 0, 1$  [130] and  $-2$ , the assumption that the critical point is at  $x = x_c$ , as given by Lemma 8.2. To see how strong the inequality is, one must substitute the conjectured exponent values. For the  $O(n = -2)$  model, the inequality is an equality. As  $n$  increases, the bound gets progressively weaker. For the  $O(n = -1)$  model, the l.h.s. of (8.23) is  $67/64$ , for the  $O(n = 0)$  model, the l.h.s. is  $73/64$ , while for the  $O(n = 1)$  model it is  $11/8$ . The exponents do not exist for the  $O(2)$  model.

As we have shown that  $\gamma_2(\alpha) - \gamma_{21}(\alpha)$  is independent of  $\alpha$ , it follows that the above inequality can be written more generally as

$$\gamma_2(\alpha) - \gamma_{21}(\alpha) \geq 1. \quad (8.24)$$

## Conjectures

### Winding angle distribution from conformal field theory

Duplantier and Saleur [81] conjectured the winding angle distribution function for the general  $O(n)$  model on a cylinder. We use the parametrisation  $n = -2 \cos(4\pi/\kappa)$  whereas Duplantier and Saleur use the symbol  $g$  which is related to  $\kappa$  by  $g = 4/\kappa$  (it is widely believed that  $\kappa = 8/3$  for SAWs). The parafermionic spin  $\sigma$  is related to  $\kappa$  by  $\sigma = \frac{3}{\kappa} - \frac{1}{2}$ . They conjecture, from CFT and Coulomb gas arguments, that

$$P(x = \theta) \propto \exp\left(-\frac{\theta^2}{2\kappa\nu \log \ell}\right), \quad \ell \rightarrow \infty, \quad (8.25)$$

where  $\ell$  is the length of the walk. Here  $\nu$  is the standard critical exponent relating the circumference of the cylinder to the length of the walk,  $L \sim \ell^\nu$ , and is given by

$$\nu = \frac{1}{4 - \kappa},$$

where  $\kappa = 2, 12/5, 8/3, 3, 4$  for  $n = -2, -1, 0, 1, 2$  respectively. Hence  $\nu$  is not defined for  $n = 2$ . Note that the winding angle distribution (8.25) is expressed entirely in terms of bulk critical exponents.

We recall from the previous sections that

$$\sum_{\theta} e^{i\tilde{\sigma}\theta} P_{\alpha}(\theta, \ell) \approx \int_{-\infty}^{\infty} e^{i\tilde{\sigma}\theta} P_{\alpha}(\theta, \ell) d\theta \propto \ell^{-\omega},$$

where  $\omega = -\gamma_{21}(\alpha) + \gamma_2(\alpha) - 1$ . The half plane corresponds to  $\alpha = \pi$ . To compare with the exponent characterising the winding angle distribution in the half plane given in Proposition 8.7 above, or more generally for a wedge with angle  $\alpha$  as given in (8.22), we thus require

$$\int_{-\infty}^{\infty} e^{\tilde{\sigma}i\theta} P(\theta) d\theta \propto \ell^{-\omega}. \quad (8.26)$$

Using (8.25) this is a straightforward integral, and gives

$$\omega = \nu \kappa \tilde{\sigma}^2 / 2 = \frac{\kappa \tilde{\sigma}^2}{2(4 - \kappa)}.$$

Hence we find

$$-\gamma_{21}(\alpha) + \gamma_2(\alpha) - 1 = \omega = \frac{9(2 - \kappa)^2}{8\kappa(4 - \kappa)}. \quad (8.27)$$

In particular we note that  $\omega = -\gamma_{21}(\alpha) + \gamma_2(\alpha) - 1$  is independent of the wedge angle  $\alpha$ . This is confirmed in the case of SAW ( $n = 0$ ) by the results of [127], and for the Ising case ( $n = 1$ ) by the results of [13].

From the existing physics literature [54, 162, 61] one can independently derive the expected values of the boundary exponents:

$$\gamma_1 = \frac{\kappa^2 + 12\kappa - 12}{8\kappa(4 - \kappa)}, \quad \gamma_{11} = -\frac{2(3 - \kappa)}{\kappa(4 - \kappa)}, \quad (8.28)$$

and thus

$$-\gamma_{11} + \gamma_1 - 1 = \frac{9(2 - \kappa)^2}{8\kappa(4 - \kappa)},$$

in perfect agreement with (8.27) for  $\alpha = \pi$ .

## Wedge exponents

The expected values of the wedge exponents have not previously been derived. However by extrapolating certain special cases we can do so. Following the notation of [127], we write the free energy of a  $d$ -dimensional wedge-shaped system as

$$F = V f_b + A f_s + L f_e + \dots,$$

where  $V$  is the  $d$ -dimensional ‘volume’ of the system,  $A$  is the  $(d - 1)$ -dimensional ‘area’ of a surface, and  $L$  is the  $(d - 2)$ -dimensional ‘length’ of the edge formed by the intersection of the two surfaces. In our case,  $d = 2$ , so the ‘volume’ is an area, the ‘area’ is the length of the boundary, and the ‘length’ corresponds to the point at the apex of a wedge. Using the scaling hypothesis, the singular part of the corresponding free energies can be written as

$$\begin{aligned} f_b &\sim t^{2-\alpha} \psi_b(ht^{-y_0\nu}) \\ f_s &\sim t^{2-\alpha_s} \psi_s(ht^{-y_0\nu}, h_1 t^{-y_1\nu}) \\ f_e &\sim t^{2-\alpha_e} \psi_e(ht^{-y_0\nu}, h_1 t^{-y_1\nu}, h_2 t^{-y_2\nu}), \end{aligned} \quad (8.29)$$

where  $t$  is the reduced temperature  $(T - T_c)/T_c$ ;  $y_0$ ,  $y_1$  and  $y_2$  are the bulk, surface and edge scaling indices, from which all the susceptibility critical exponents follow. The reduced magnetic fields in the bulk, surface and edge are denoted, respectively  $h$ ,  $h_1$  and  $h_2$ . In particular, the bulk susceptibility is given by

$$\chi = \partial^2 f_b / \partial h^2 \asymp t^{-\gamma}; \quad \gamma = \nu(2y_0 - d),$$

the surface susceptibilities are given by

$$\chi_1 = \partial^2 f_s / \partial h \partial h_1 \asymp t^{-\gamma_1}, \quad \gamma_1 = \nu(y_0 + y_1 - d + 1),$$

$$\chi_{11} = \partial^2 f_s / \partial h_1^2 \asymp t^{-\gamma_{11}}, \quad \gamma_{11} = \nu(2y_1 - d + 1),$$

and the edge susceptibilities are given by

$$\chi_2 = \partial^2 f_e / \partial h \partial h_2 \asymp t^{-\gamma_2}, \quad \gamma_2 = \nu(y_0 + y_2 - d + 2),$$

$$\chi_{21} = \partial^2 f_e / \partial h_1 \partial h_2 \asymp t^{-\gamma_{21}}, \quad \gamma_{21} = \nu(y_1 + y_2 - d + 2).$$

In [127] it was shown (non-rigorously) that, for the  $O(n = 0)$  model,  $y_2(\alpha) = -5\pi/8\alpha$ , where  $\alpha$  is the wedge angle. Similarly, in [13] it was shown (non-rigorously) that, for the  $O(n = 1)$  model,  $y_2(\alpha) = -\pi/2\alpha$ . For the Gaussian model, corresponding to the  $O(n = -2)$  model,  $y_2(\alpha) = -\pi/\alpha$ . Thus for these three cases, we have

$$y_2(\alpha) = -\pi\sigma/\alpha.$$

If this is true for other values of  $n$ , we then find

$$\gamma_2 = \frac{\kappa^2 + 8\kappa + 12 - 24\pi/\alpha + 4\pi\kappa/\alpha}{8\kappa(4 - \kappa)},$$

and

$$\gamma_{21} = \frac{3\kappa - 6 - 6\pi/\alpha + \pi\kappa/\alpha}{2\kappa(4 - \kappa)}.$$

where we have used  $\sigma = \frac{3}{\kappa} - \frac{1}{2}$ . If we set  $\alpha = \pi$ , these equations reduce to (8.28), providing evidence for the validity of our assumption that these equations hold for all  $n \in [-2, 2)$ . For  $n = 2$  the free energy is believed to exhibit an essential singularity, so that critical exponents do not exist. This is signalled in the conjectured exponent values by the divergence at  $\kappa = 4$ , corresponding to the  $O(n = 2)$  model.

We remark that the scaling indices  $y_0$ ,  $y_1$  and  $y_2$  take particularly simple forms if parameterised in terms of  $\sigma$ :

$$y_0 = \frac{(\sigma + 1)(\sigma + 2)}{2\sigma + 1}, \quad y_1 = 1 - \sigma, \quad y_2 = -\frac{\pi\sigma}{\alpha}.$$

## Conclusion

We have generalised the identity of Duminil-Copin and Smirnov off criticality, which allows us to make statements about critical exponents. We have proved an inequality for surface and wedge exponents, subject only to their existence.

We have similarly proved, under the same assumption, a relationship between the surface susceptibility exponents and the winding angle exponent of the  $O(n)$  model for  $n \in [-2, 2)$ . Previously conjectured values of the surface and winding angle exponents are in agreement with the relationship we have derived for all values of  $n \in [-2, 2)$ .

A study of the edge exponents that arise when considering the  $O(n)$  model in a wedge geometry permits us to conjecture the exact value of the exponents for all wedge angles.



# Bibliography

- [1] N I Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis*, London: Oliver and Boyd (1965)
- [2] M Albert, M Atkinson and S Linton, *Permutations generated by stacks and dequeues*, Ann. Comb. **14** (2010) 3–16.
- [3] M Albert and M Bousquet-Mélou, *Permutations sortable by two stacks in parallel and quarter-plane walks*, Europ. J. Comb. **43** (2015), 131–164.
- [4] M Albert, M Elder, A Rechnitzer, P Westcott and M Zabrocki, *On the Stanley-Wilf limit of 4231-avoiding permutations and a conjecture of Arratia*, Adv. in Appl. Math. **36**(2) (2006), 95–105.
- [5] M Albert, J Pantone, V Vatter, *On the growth of merges and staircases of permutation classes*, arXiv:1608.06969 (2016).
- [6] R E Aldred, M D Atkinson, H P van Ditmarsch, C C Handley, D A Holton and D J McCaughan, *Permuting machines and priority queues*, Theoretical Computer Science, 349(3), 309–317 (2005).
- [7] J. Ambjørn, T. G. Budd, *Trees and spatial topology change in causal dynamical triangulations*, Journal of Physics A: Mathematical and Theoretical, **46**(31), (2013) 315201.
- [8] O Angel and O Schramm. *Uniform infinite planar triangulations*, Comm. Math. Phys., **241**(2-3) (2003), 191–213.
- [9] R Arratia, *On the Stanley-Wilf conjecture for the number of permutations avoiding a given pattern*, Electron. J. Combin. **6** (1999).
- [10] M Atapour and N Madras, *Large deviations and ratio limit theorems for pattern-avoiding permutations*, Combinatorics, Probability and Computing **23**(2) (2014) 161–200.

- [11] M D Atkinson, M M Murphy and N Ruškuc, *Sorting with two ordered stacks in series*, Theor. Comp. Sci. **289** (2002) 205–323.
- [12] C Banderier and P Flajolet, *Basic analytic combinatorics of directed lattice paths*, Theor. Comp. Sci. **281**(1-2) (2002) 37–80.
- [13] M N Barber, I Peschel and P A Pearce, Magnetization at corners in two-dimensional Ising models, *J. Stat. Phys.* **37** (1984), 497—527.
- [14] J F Barbero Gonzalez, J Salas and E J S Villaseñor, *Bivariate generating functions for a class of linear recurrences: General structure*, J. Combin. Theory A **125** (2014) 146–165.
- [15] J F Barbero Gonzalez, J Salas and E J S Villaseñor, *Generalized Stirling permutations and forests: higher-order Eulerian and Ward numbers*, Electron. J. Combin. **22**(3) (2015) paper 3.37.
- [16] P Barry, *Continued fractions and transformations of integer sequences*, J. Integer Seq. **12** (2009), article 09.7.6.
- [17] R J Baxter, *Dichromatic polynomials and Potts models summed over rooted maps*, Ann. Comb., **5**(1) (2002), 17–36.
- [18] N. Beaton, M. Bousquet-Mélou, J. de Gier and A.J. Guttmann, The critical fugacity for surface adsorption of SAW on the honeycomb lattice is  $1 + \sqrt{2}$ , *arXiv:1109.0358v3*.
- [19] J Belk and K S Brown, *Forest diagrams for elements of Thompson’s group  $F$* , Int. J. Alg. Comp. **15** (2005) 815–850 .
- [20] J Bell and M Mishna, *On the Complexity of the Cogrowth Sequence*, *arXiv:1805.08118* (2018).
- [21] O Bernardi, *Tutte polynomial, subgraphs, orientations and sandpile model: new connections via embeddings*, Electron. J. Combin., **15**(1) (2008) Research Paper 109, 53.
- [22] O Bernardi and N Bonichon, *Intervals in Catalan lattices and realizers of triangulations*, J. Combin. Theory Ser. A, **116**(1) (2009), 55–75.
- [23] O Bernardi and M Bousquet-Mélou, *Counting colored planar maps: algebraicity results*, J. Combin. Theory Ser. B, **101**(5) (2011), 315–377.
- [24] O Bernardi and M Bousquet-Mélou, *Counting coloured planar maps: differential equations*, Comm. Math. Phys., **354**(1) (2017), 31–84.

- [25] O Bernardi, M Bousquet-Mélou and K Raschel, Counting quadrant walks via Tutte’s invariant method. ArXiv:1708.08215 (2017).
- [26] D Bessis, C Itzykson and J B Zuber, *Quantum field theory techniques in graphical enumeration*, Adv. in Appl. Math., **1**(2) (1980), 109–157.
- [27] D Bevan, *Permutations avoiding 1324 and patterns in Łukasiewicz paths*, J. London Math. Soc. **92**(1) (2015), 105–122.
- [28] D Bevan, R Brignall, A Elvey Price and J Pantone, *New bounds on the growth rate of 1324-avoiders*, arXiv:1711.10325 (2017)
- [29] K Binder and P C Hohenberg, *Phase transitions and static spin correlations in Ising models with free surfaces*, Phys. Rev. B **6** (1972), 3461–3487.
- [30] C Bleak, M G Brin and J T Moore, *Complexity among the finitely generated subgroups of Thompson’s group*, arXiv:1711.10998 (2017).
- [31] N Blitvić, *The  $(q, t)$ -Gaussian process*, J. Funct. Anal. **263** (2012) 3270–3305.
- [32] B Bollobás, *Modern graph theory, volume 184 of Graduate Texts in Mathematics*, Springer-Verlag, New York, (1998).
- [33] M Bóna, *Exact enumeration of 1342-avoiding permutations: a close link with labeled trees and planar maps*, J. Combin. Theory Ser. A, **80**(2) (1997), 257–272.
- [34] M Bóna, *A survey of stack-sorting disciplines*, the electronic journal of combinatorics **9**(2) (2003).
- [35] M Bóna, *A simple proof for the exponential upper bound for some tenacious patterns*, Adv. in Appl. Math. **33**(1) (2004) 192–198.
- [36] M Bóna, *The limit of a Stanley-Wilf sequence is not always rational, and layered patterns beat monotone patterns*, J. Comb. Theory Ser. A **110**(2) (2005) 223–235.
- [37] M Bóna, *New records in Stanley-Wilf limits*, Europ. J. Comb. **28**(1) (2007) 75–85.
- [38] M Bóna, *Combinatorics of Permutations, second ed. Discrete Mathematics and its Applications*, CRC Press, (2012).

- [39] M Bóna, *A new upper bound for 1324-avoiding permutations*, Combin. Probab. Comput. **23**(5) (2014), 717–724.
- [40] M Bóna, *A new record for 1324-avoiding permutations*, Eur. J. Math. **1**, 1 (2015), 198–206.
- [41] N Bonichon, *A bijection between realizers of maximal plane graphs and pairs of non-crossing Dyck paths*, Discrete Math, **298**(1-3) (2005) 104–114.
- [42] N Bonichon, M Bousquet-Mélou, P Dorbec and C Pennarun, *On the number of planar Eulerian orientations*, Europ. J. Combinatorics **65** (2017) 59–91.
- [43] N Bonichon, M Bousquet-Mélou and É Fusy, *Baxter permutations and plane bipolar orientations*, Sémin. Lothar. Combin., 61A:Art. B61Ah, 29, (2010).
- [44] G Borot, J Bouttier and B Duplantier, *Nesting statistics in the  $O(n)$  loop model on random planar maps*, arXiv:1605.02239 (2016).
- [45] G Borot, J Bouttier and E Guitter, *Loop models on random maps via nested loops: case of domain symmetry breaking and application to the Potts model*, J. Phys. A, **45** (2012) 494017.
- [46] M Bousquet-Mélou, *Four classes of pattern-avoiding permutations under one roof: generating trees with two labels*. Electron. J. Combin., 9(2):Research paper 19, 31 pp. (electronic), 2002/03. Permutation patterns (Otago, 2003).
- [47] M Bousquet-Mélou and S Butler, *Forest-like permutations*, Ann. Comb. **11**(3-4) (2007), 335–354.
- [48] M Bousquet-Mélou and J Courtiel, *Spanning forests in regular planar maps*, J. Combin. Theory Ser. A, **135** (2015) 1–59.
- [49] M Bousquet-Mélou and A Elvey Price, *The generating function of planar Eulerian orientations* arXiv:1803.08265 (2018).
- [50] M Bousquet-Mélou, É Fusy and K Raschel, *Bipolar orientations and quadrant walks*. In preparation.
- [51] M Bousquet-Mélou, A Jehanne, *Polynomial equations with one catalytic variable, algebraic series and map enumeration*, J. Comb. Theory Ser. B **96**(5) (2006), 623–672.

- [52] J Bouttier, P Di Francesco and E Guitter, *Planar maps as labeled mobiles*, Electron. J. Combin. **11**(1) (2004), Research Paper 69, 27 pp. (electronic).
- [53] J Bouttier, É Fusy and E Guitter, *On the two-point function of general planar maps and hypermaps*, arXiv:1312.0502 (2013).
- [54] A J Bray and M A Moore, *Surface critical exponents in terms of bulk exponents*, Phys. Rev. Lett. **38** (1977), 1046–1048.
- [55] E Brézin, C Itzykson, G Parisi, and J B Zuber, *Planar diagrams*, Comm. Math. Phys., **59**(1) (1978), 35–51.
- [56] M G Brin, *Elementary amenable subgroups of R. Thompson’s group F*, Intl. J. Alg. and Comp. **15** (2005), 619–642.
- [57] W G Brown, *Enumeration of non-separable planar maps*, Canad. J. Math. **15** (1963), 526–545.
- [58] D Callan, *A combinatorial survey of identities for the double factorial*, preprint (2009), arXiv:0906.1317.
- [59] D Callan, T Mansour and M Shattuck, *Some identities for derangement and Ward number sequences and related bijections*, Pure Math. Appl. (P.U.M.A.) **25** (2015) 132–143.
- [60] J W Cannon, W J Floyd and W R Parry. *Introductory notes on Richard Thompson’s groups*, Enseignement Mathématique **42** (1996), 215–256.
- [61] J. Cardy, *Conformal invariance and surface critical behavior*, Nucl. Phys. B, **240** (1984), 514–532.
- [62] L Carlitz, *Note on the numbers of Jordan and Ward*, Duke Math. J. **38** (1971) 783–790.
- [63] P Chassaing and B Durhuus, *Local limit of labeled trees and expected volume growth in a random quadrangulation*, Ann. Probab., **34**(3) (2006), 879–917.
- [64] P Chassaing and G Schaeffer, *Random planar lattices and integrated superBrownian excursion*, Probab. Theory Related Fields, **128**(2) (2004), 161–212.
- [65] A Claesson, V Jelínek and E Steingrímsson, *Upper bounds for the Stanley-Wilf limit of 1324 and other layered patterns*, J. Comb. Theory Ser. A **119**(8) (2012), 1680–1691.

- [66] L. Clark, *Asymptotic normality of the Ward numbers*, Discrete Math. **203** (1999) 41–48.
- [67] J M Cohen, *Cogrowth and amenability of discrete groups*, J. Funct. Anal, **48** (1982), 301–309.
- [68] L Comtet, *Sur le quatrième problème et les nombres de Schröder*, Comptes Rendus Acad. Sci. Paris Sér. A-B **271** (1970) A913–A916.
- [69] L Comtet, *Advanced Combinatorics: The Art of Finite and Infinite Expansions* (Reidel, Dordrecht–Boston, 1974). [French original: *Analyse Combinatoire*, tomes I et II, Presses Universitaires de France, Paris, 1970.]
- [70] A R Conway and A J Guttmann, *On 1324-avoiding permutations*, Adv. in Appl. Math. **64** (2015), 50–69.
- [71] A R Conway, A J Guttmann and P Zinn-Justin, *1324-avoiding permutations revisited*, Adv. in Appl. Math. **96** (2018) 312–333.
- [72] N Curien and J-F Le Gall, *The Brownian plane*, J. Theoret. Probab., **27**(4) (2014), 1249–1291.
- [73] N Curien and G Miermont, *Uniform infinite planar quadrangulations with a boundary*, Random Structures Algorithms, **47**(1) (2015), 30–58.
- [74] D. Denton and P. Doyle, [oeis.org/A182216/b182216.txt](http://oeis.org/A182216/b182216.txt)
- [75] P W Diaconis and S P Holmes, *Matchings and phylogenetic trees*, Proc. Natl. Acad. Sci. USA **95** (1998) 14600–14602.
- [76] E Domany, D Mukamel, B Nienhuis and A Schwimmer, *Duality relations and equivalences for models with  $O(N)$  and cubic symmetry*, Nucl. Phys. B **190** 279–287 (1981).
- [77] E Duchi, V Guerrini, S Rinaldi and G Schaeffer, *Fighting fish*, J. Phys. A: Math. Theor. **50**(2) (2017), 024002.
- [78] E Duchi, V Guerrini, S Rinaldi and G Schaeffer, *Fighting fish: enumerative properties*, Sémin. Lothar. Combin. 78B (2017), Art. 43, 12 pp.
- [79] H. Duminil-Copin and S. Smirnov, *The connective constant of the honeycomb lattice equals  $\sqrt{2 + \sqrt{2}}$* , Annals of Mathematics (2012) 1653–1665.
- [80] B Duplantier and I Kostov, *Conformal spectra of polymers on a random surface*, Phys. Rev. Lett., **61**(13) (1988), 1433–1437.

- [81] B Duplantier and H Saleur, *Winding-angle distributions of two-dimensional self-avoiding walks from conformal invariance*, Phys. Rev. Lett. **60** (1988), 2343–2346.
- [82] G Edgar *Is the sequence of Apéry numbers a Stieltjes moment sequence?*, 22 August 2014, <http://mathoverflow.net/questions/179108/>.
- [83] G Edgar *The Apéry numbers as a Stieltjes moment sequence*, in preparation.
- [84] E S Egge, *Defying God: the Stanley-Wilf conjecture, Stanley-Wilf limits, and a two generation explosion of combinatorics*, A century of advancing mathematics. Math. Assoc. America, Washington, DC, (2015), 65–82.
- [85] M Elder, É Fusy and A R Rechnitzer, *Counting elements and geodesics in Thompson’s group  $F$* , J. Alg. **324** (2010), 102–121.
- [86] M Elder, G Lee and A R Rechnitzer, *Permutations generated by a depth 2 stack and an infinite stack in series are algebraic.*, EJC, **22**(2) (2015), 2–16.
- [87] M Elder, A Rechnitzer and E J Janse van Rensburg, *Random sampling of trivial words in finitely presented groups*, Experimental Math. **24**(4) (2015), 391–409.
- [88] M Elder, A Rechnitzer, E J Janse van Rensburg and T Wong *The cogrowth series for  $BS(N, N)$  is  $D$ -finite*, Internat. J. Alg. Comp. **24**(02) (2014), 171–187.
- [89] M Elder, A R Rechnitzer and T Wong, *On the cogrowth of Thompson’s group  $F$* , Groups - Complexity - Cryptology, **4** (2012), 301–320.
- [90] A Elvey Price, J de Gier, A J Guttmann and A Lee, *Off-critical parafermions and the winding angle distribution of the  $O(n)$  model*, J Phys. A: Math. Gen. **45**(27) (2012), 275002.
- [91] A Elvey Price and A J Guttmann, *Counting planar Eulerian orientations*, Europ. J. Combinatorics **71** (2018), 73–98.
- [92] A Elvey Price and A J Guttmann, *Numerical studies of Thompson’s group  $F$  and related groups*, to appear in Int. J. Alg. and Comp.
- [93] A Elvey Price and A J Guttmann, *Permutations sortable by dequeues and by two stacks in parallel*, Europ. J. Combinatorics **59** (2017), 71–95.

- [94] A Elvey Price and A J Guttmann, *Permutations sortable by two stacks in series*. Advances in Applied Mathematics **83** (2017), 81–96.
- [95] A Elvey Price, M A Surani and S Zhou, *The isoperimetric number of the incidence graph of  $PG(n, q)$* , The Electronic Journal of Combinatorics **25**(3) (2018).
- [96] G Eneström, *Die Schriften Eulers chronologisch nach den Jahren geordnet, in denen sie verfaßt worden sind*, Jahresbericht der Deutschen Mathematiker-Vereinigung (Teubner, Leipzig, 1913).
- [97] L. Euler, *Introductio in Analysin Infinitorum*, tomus primus (Bousquet, Lausanne, 1748). English translation: L Euler, *Introduction to Analysis of the Infinite, Book I*, translated from the Latin and with an introduction by John D. Blanton (Springer-Verlag, New York, 1988). [Latin original and French and German translations available at <http://eulerarchive.maa.org/pages/E101.html>]
- [98] L Euler, *De seriebus divergentibus*, Novi Commentarii Academiae Scientiarum Petropolitanae **5**, 205–237 (1760). [Latin original and English and German translations available at <http://eulerarchive.maa.org/pages/E247.html>]
- [99] L Euler, *De transformatione seriei divergentis  $1 - mx + m(m+n)x^2 - m(m+n)(m+2n)x^3 + \text{etc. in fractionem continuam}$* , Nova Acta Academiae Scientiarum Imperialis Petropolitanae **2**, 36–45 (1788). [Latin original and English and German translations available at <http://eulerarchive.maa.org/pages/E616.html>]
- [100] S. Even and A. Itai, *Queues, stacks and graphs*, Theory of Machines and Computations Proc. Internat. Sympos., Technion, Haifa, 1971, Academic Press, New York, (1971), 71–86.
- [101] W Fang, *Fighting fish and two-stack sortable permutations*, arXiv:1711.05713 (2017)
- [102] S Felsner, É Fusy, M Noy, and D Orden, *Bijections for Baxter families and related objects*, J. Combin. Theory Ser. A, **118**(3) (2011) 993–1020.
- [103] P Flajolet, *Analytic models and ambiguity of context-free languages*, Theoret. Comput. Sci., **49**(2-3) (1987) 283–309.
- [104] P Flajolet, *Combinatorial aspects of continued fractions*, Discrete Math. **32** (1980) 125–161.



- [105] P Flajolet and A Odlyzko, *Singularity analysis of generating functions*, SIAM J. Discrete Math., **3**(2) (1990) 216–240.
- [106] P Flajolet and R Sedgewick, *Analytic Combinatorics*. Cambridge University Press, Cambridge (2009).
- [107] P Flajolet, P Zimmerman and B Van Cutsem, *A calculus for the random generation of labelled combinatorial structures*, Theoret. Comput. Sci. **132** (1994) 1–35.
- [108] S B Fordham, *Minimal length elements of Thompsons group  $F$* , Geom. Dedicata **99** (2003), 179–220.
- [109] J Fox, *Stanley-Wilf limits are typically exponential*, arXiv:1310.8378 (2013), to appear in Adv. Math.
- [110] É Fusy and E Guitter, *Comparing two statistical ensembles of quadrangulations: A continued fraction approach*, Ann. Inst. Henri Poincaré D **4** (2017) 125–176.
- [111] É Fusy, D Poulalhon, and G Schaeffer, *Bijective counting of plane bipolar orientations and Schnyder woods*, European J. Combin., **30**(7) (2009) 1646–1658.
- [112] F Gantmakher and M Krein, *Sur les matrices complètement non négatives et oscillatoires*, Compositio Mathematica **4** (1937), 445–476.
- [113] S Garrabrant and I Pak, *Words in linear groups, random walks, automata and  $P$ -recursiveness*, J. Comb. Algebra **1**(2) (2017), 127–144.
- [114] P-G de Gennes, *Exponents for the excluded-volume problem as derived by the Wilson method*, Phys. Lett. A, **38** (1972), 339–340.
- [115] I M Gessel, *Symmetric functions and  $P$ -recursiveness*, J. Comb. Theory Ser. A **53**(2) (1990), 257–285.
- [116] I Gessel and R P Stanley, *Stirling polynomials*, J. Combin. Theory A **24** (1978) 24–33.
- [117] S.N. Gladkovskii, 1 May 2013, contribution to [198, sequence A000311].
- [118] R L Graham, D E Knuth and O Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. (Addison-Wesley, Reading, Mass., 1994).

- [119] D Gretete, *Random walks on a discrete Heisenberg group*, Rend. Circ. Mat. Palermo **60** (2011) 329–335.
- [120] R I Grigorchuk, *Symmetric random walks on discrete groups*, Multicomponent random systems (eds. R L Dobrushin and Ya. G Sinai), Nauka, Moscow (1978).
- [121] V S Guba, *On the properties of the Cayley graph of Richard Thompson’s group  $F$* , Internat. J. Alg. Comp. **14** (5-6) 677–702, International Conference on Semigroups and Groups in honor of the 65th birthday of Prof. John Rhodes (2004).
- [122] A J Guttmann, *Phase Transitions and Critical Phenomena*, vol 13, eds. C Domb and J Lebowitz, Academic Press, London and New York, (1989).
- [123] A J Guttmann, *Analysis of series expansions for non-algebraic singularities*, J. Phys A: Math. Theor. **48**(4) (2015), 045209.
- [124] A J Guttmann, *Series extension: Predicting approximate series coefficients from a finite number of exact coefficients*, J. Phys A: Math. Theor. **49**(41) (2016), 415002.
- [125] A J Guttmann and I Jensen, *Series Analysis*, Chapter 8 of *Polygons, Polyominoes and Polycubes* Lecture Notes in Physics **775**, ed. A J Guttmann, Springer, (Heidelberg), (2009).
- [126] A J Guttmann and G S Joyce, *A new method of series analysis in lattice statistics*, J Phys A, **5** (1972), 81–84.
- [127] A J Guttmann and G M Torrie, *Critical behaviour at an edge for the SAW and Ising model*, J. Phys. A **17** (1984), 3539–3552.
- [128] S Haagerup, U Haagerup and M Ramirez-Solano, *A computational approach to the Thompson group  $F$* , Int. J. Alg. and Comp. **25** (2015), 381–432.
- [129] J M Hammersley and S G Whittington, *Self-avoiding walks in wedges*, J. Phys. A **17** (1985), 101–111.
- [130] R M F Houtappel, *Order-disorder in hexagonal lattices*, Physica **16** (1950), 425–455

- [131] D M Jackson, *Counting cycles in permutations by group characters, with application to a topological problem*, Trans. Amer. Math. Soc., **299**(2) (1987), 785–801.
- [132] D M Jackson and T I Visentin, *A character-theoretic approach to embeddings of rooted maps in an orientable surface of given genus*, Trans. Amer. Math. Soc., **322**(1) (1990), 343–363.
- [133] M. Josuat-Vergès, *A  $q$ -analog of Schläfli and Gould identities on Stirling numbers*, Ramanujan J. **46** (2018) 483–507.
- [134] A Kasraoui and J Zeng, *Distribution of crossings, nestings and alignments of two edges in matchings and partitions*, Electron. J. Combin. **13** (2006) #R33.
- [135] A. Kassel and D. B. Wilson, *The looping rate and sandpile density of planar graphs*, Amer. Math. Monthly, **123**(1) (2016), 19–39.
- [136] V A Kazakov, *Ising model on a dynamical planar random lattice: exact solution*, Phys. Lett. A, **119**(3) (1986) 140–144.
- [137] V A Kazakov and P Zinn-Justin, *Two-matrix model with ABAB interaction*, Nuclear Phys. B, **546**(3) (1999) 647–668.
- [138] R Kenyon, J Miller, S Sheffield and D B Wilson, *Bipolar orientations on planar maps and SLE<sub>12</sub>*, arXiv:1511.04068 (2015).
- [139] S Kitaev, *Patterns in Permutations and Words*, Springer (2011).
- [140] M Klazaar, *Bell numbers, their relatives, and algebraic differential equations*, J. Comb. Theory Ser. A **102**(1) (2003), 63–87.
- [141] D E Knuth, *Fundamental Algorithms, The Art of Computer Programming Vol 1*, Addison-Wesley, Massachusetts (1968)
- [142] D E Knuth, *Fundamental Algorithms, Vol 3*, Addison-Wesley, Massachusetts (1973)
- [143] I K Kostov, *Exact solution of the six-vertex model on a random lattice*, Nucl. Phys B **575**(3) (2000), 513–534.
- [144] B Leclerc, *Les hiérarchies de parties et leur demi-treillis*, Math. Sci. Humaines **89** (1985) 5–34, 67; errata **92**, 40 (1985).
- [145] J-F Le Gall, *The topological structure of scaling limits of large planar maps*, Invent. Math., **169**(3) (2007) 621–670.

- [146] J-F Le Gall, *Random geometry on the sphere*, In S. Y. Jang, Y. R. Kim, D.-W. Lee, and I. Yie, editors, International Congress on Mathematicians, Plenary lectures and ceremonies, volume 1 (2014) 421–442.
- [147] J-F Le Gall and G Miermont, *Scaling limits of random planar maps with large faces*, Ann. Probab., **39**(1) (2011), 1–69.
- [148] L Lipshitz, *The diagonal of a  $D$ -finite power series is  $D$ -finite*, J. Algebra, **113**(2) (1988) 373–378.
- [149] L Lipshitz,  *$D$ -finite power series*, J. Algebra, **122** (1989) 353–373.
- [150] J-G Liu and R Pego, *On generating functions of Hausdorff moment sequences*, Transactions of the American Mathematical Society **368**(12) (2016) 8499–8518.
- [151] P A MacMahon, *Combinatorial Analysis* Vols I and II originally published in 1915 and 1916. Republished as one volume in 1960 by Chelsea Publishing Co, and in 2004 by Dover NY.
- [152] N. Makarov and S. Smirnov, *Off-critical lattice models and massive SLEs*, XIVth International Congress on Mathematical Physics, Lisbon, Portugal, July 28 - August 2, 2003, 362–371. World Sci. Publ., Singapore, 2009, arXiv:0909.5377.
- [153] J-F Marckert and A Mokkadem, *Limit of normalized quadrangulations: the Brownian map*, Ann. Probab., **34**(6) (2006) 2144–2202.
- [154] A Marcus and G Tardos, *Excluded permutation matrices and the Stanley-Wilf conjecture*, J. Comb. Theory Ser. A **107**(1) (2004), 153–160.
- [155] L Ménard, *The two uniform infinite quadrangulations of the plane have the same law*, Ann. Inst. Henri Poincaré Probab. Stat., **46**(1) (2010) 190–208.
- [156] G Miermont, *Tessellations of random maps of arbitrary genus*, Annales scientifiques de l’Ecole normale supérieure, **42**(5) (2009) 725–781.
- [157] G Miermont, *The Brownian map is the scaling limit of uniform random plane quadrangulations*, Acta Math., **210**(2) (2013) 319–401.
- [158] J T Moore, *Fast growth in the Følner function for Thompson’s group  $F$* , Groups Geom. Dyn. **7** (2013) 633–651.

- [159] R C Mullin, *On the enumeration of tree-rooted maps*, Canad. J. Math., **19** (1967) 174–183.
- [160] M M Murphy, *Restricted permutations, antichains, atomic classes, stack sorting*, Ph.D. thesis, University of St Andrew (2002).
- [161] A Navas, *Quelques groupes moyennables de difféomorphismes de l'intervalle*, Bol. Soc. Mat. Mexicana **10** (2004), 219–244.
- [162] B Nienhuis, *Exact critical point and critical exponents of  $O(n)$  models in two dimensions*, Phys. Rev. Lett., **49** (1982), 1062–1065.
- [163] M Noy, *Enumeration of noncrossing trees on a circle*, Discrete Math. **180**(1) (1998), 301–313.
- [164] R Oste and J Van der Jeugt, *Motzkin paths, Motzkin polynomials and recurrence relations*, Electron. J. Combin. **22**(2) (2015) #P2.8.
- [165] I Pak, *Complexity problems in enumerative combinatorics*, arXiv:1803.06636 (2018).
- [166] M Pétréolle, A D Sokal and B-X Zhu, *Lattice paths and branched continued fractions: An infinite sequence of generalizations of the Stieltjes–Rogers and Thron–Rogers polynomials, with coefficientwise Hankel-total positivity*, preprint (2018), arXiv:1807.03271.
- [167] A Pierrot, *Combinatoire et algorithmique dans les classes de permutations*, PhD thesis (2013)
- [168] A Pierrot and D Rossin, *2-stack sorting is polynomial*, arXiv:1304.2860v1 (2013).
- [169] C Pittet and L Saloff-Coste, *On the stability of the behavior of random walks on groups*, J. Geom. Anal. **10**(4) (2000), 713–737.
- [170] C Pittet and L Saloff-Coste, *On random walks on wreath products*, Ann. of Probab. **30**(2) (2002), 948–977.
- [171] V R Pratt, *Computing permutations with double-ended queues, parallel stacks and parallel queues*, in STO '73: Proceedings of the Fifth Annual ACM Symposium on Theory of Computing, ACM Press, New York, NY, USA (1973) 268–277.
- [172] H Prodinger, *A bijection between phylogenetic trees and plane oriented recursive trees*, preprint (2017), arXiv:1709.05966.

- [173] E M Rains, *Increasing subsequences and the classical groups*, The Electronic Journal of Combinatorics **5**(1) (1998), 12.
- [174] D Revelle, *Heat kernel asymptotics on the lamplighter group*, Elect. Comm. in Probab. **8** (2003), 142–154.
- [175] L Richier, *Universal aspects of critical percolation on random half-planar maps*, Electron. J. Probab., **20** (2015) Paper No. 129, 45.
- [176] J Riordan, *Combinatorial Identities* (Wiley, New York, 1968). [Reprinted with corrections by Robert E. Krieger Publishing Co., Huntington NY, 1979.]
- [177] J Riordan, *The blossoming of Schröder’s fourth problem*, Acta Math. **137** (1976) 1–16.
- [178] E Roblet and X G Viennot, *Théorie combinatoire des T-fractions et approximants de Padé en deux points*, Discrete Math. **153** (1996) 271–288.
- [179] G Schaeffer, *Bijective census and random generation of Eulerian planar maps with prescribed vertex degrees*, Electron. J. Combin., **4**(1) (1997) Research Paper 20, 14 pp. (electronic).
- [180] G Schaeffer, *Conjugaison d’arbres et cartes combinatoires aléatoires*, PhD thesis, Université Bordeaux I (1998).
- [181] E Schröder, *Vier kombinatorische Probleme*, Z. für Math. Physik **15** (1870) 361–376.
- [182] S Sheffield, *Quantum gravity and inventory accumulation*, Ann. Probab., **44**(6) (2016) 3804–3848.
- [183] R Simion, F W Schmidt, *Restricted permutations*, Europ. J. Combinatorics, **6** (1985), 383–406.
- [184] M F Singer, *Algebraic relations among solutions of linear differential equations*, Trans. Amer. Math. Soc., **295**(2) (1986) 753–763.
- [185] S Smirnov, *Discrete Complex Analysis and Probability*, Proceedings of the International Congress of Mathematicians (ICM), Hyderabad, India, arXiv:1009.6077 (2010).

- [186] A D Sokal, *Coefficientwise total positivity (via continued fractions) for some Hankel matrices of combinatorial polynomials*, talk at the Séminaire de Combinatoire Philippe Flajolet, Institut Henri Poincaré, Paris, 5 June 2014; transparencies available at <http://semflajolet.math.cnrs.fr/index.php/Main/2013-2014>
- [187] A D Sokal, *The Euler and Springer numbers as moment sequences*, preprint (2018), arXiv:1804.04498.
- [188] A D Sokal, *Coefficientwise total positivity (via continued fractions) for some Hankel matrices of combinatorial polynomials*, in preparation.
- [189] A D Sokal and J Zeng, *Some multivariate master polynomials for permutations, set partitions, and perfect matchings, and their continued fractions*, in preparation.
- [190] Z E Stankova, *Forbidden subsequences*, Discrete Math, **132**(1-3) (1994), 291–316.
- [191] Z E Stankova, *Classification of forbidden subsequences of length 4*, Europ. J. Combinatorics, **17**(5) (1996), 501–517.
- [192] H E Stanley, *Dependence of Critical Properties on Dimensionality of Spins*, Phys. Rev. Lett., **20** (1968) 589–593.
- [193] R P Stanley, *Enumerative Combinatorics*, vol. 2 (Cambridge University Press, Cambridge–New York, 1999).
- [194] R P Stanley, *Catalan Numbers*, Cambridge Univ. Press, (2015).
- [195] E Steingrímsson, *Some open problems on permutation patterns*, Surveys in Combinatorics 2013, vol. 409 of London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, (2013), 239–263.
- [196] T-J Stieltjes, *Recherches sur les fractions continues*, Annales de la Faculté des sciences de Toulouse: Mathématiques. **8**(4) (1894), 1–122.
- [197] R. Tarjan, *Sorting using networks of queues and stacks*, J. Assoc. Comput. Mach. **19** (1972), 341–346.
- [198] The OEIS Foundation Inc. The On-Line Encyclopedia of Integer Sequences. Published electronically at <https://oeis.org>.
- [199] W T Tutte, *A contribution to the theory of chromatic polynomials*, Canadian J. Math., **6** (1954), 80–91.

- [200] W T Tutte, *A census of planar triangulations*, Canad. J. Math., **14** (1962), 21–38.
- [201] W T Tutte, *A census of planar maps*, Canad. J. Math., **15** (1963), 249–271.
- [202] W T Tutte, *Dichromatic sums for rooted planar maps*, In Combinatorics (Proc. Sympos. Pure Math., Vol. XIX, Univ. California, Los Angeles, Calif., 1968) (1971), 235–245.
- [203] W T Tutte, *Chromatic sums for rooted planar triangulations: the cases  $\lambda = 1$  and  $\lambda = 2$* , Canad. J. Math., **25** (1973), 426–447.
- [204] W Unger, *On the  $k$ -colouring of circle-graphs*, Annual Symposium on Theoretical Aspects of Computer Science (1988), Springer, Berlin, Heidelberg.
- [205] W Unger, *The complexity of colouring circle graphs*, Annual Symposium on Theoretical Aspects of Computer Science (1992), Springer, Berlin, Heidelberg.
- [206] V Vatter, *Permutation classes*, The Handbook of Enumerative Combinatorics, M. Bóna, Ed. CRC Press (2015), 753–833.
- [207] G. Viennot, *Une théorie combinatoire des polynômes orthogonaux généraux*, Notes de conférences données à l’Université du Québec à Montréal, septembre-octobre 1983. Available on-line at [http://www.xavierviennot.org/xavier/polynomes\\_orthogonaux.html](http://www.xavierviennot.org/xavier/polynomes_orthogonaux.html)
- [208] H S Wall, *Analytic Theory of Continued Fractions* (Van Nostrand, New York, 1948).
- [209] M Ward, *The representation of Stirling’s numbers and Stirling’s polynomials as sums of factorials*, Amer. J. Math. **56** (1934) 87–95.
- [210] D Welsh, *The Tutte polynomial*, Random Structures Algorithms, **15**(3-4) (1999), 210–228. Statistical physics methods in discrete probability, combinatorics, and theoretical computer science (Princeton, NJ, 1997).
- [211] J West, *Sorting twice through a stack*, Theor. Comp. Sci. **117** (1993) 303–313.
- [212] Wolfram Research, Inc. Mathematica. Version 10.3. [www.wolfram.com/mathematica](http://www.wolfram.com/mathematica), (2015).



- [213] D Zeilberger, *A proof of Julian West's conjecture that the number of two-stack-sortable permutations of length  $n$  is  $2(3n)!/((n+1)!(2n+1)!)$ .* Discrete Math. **102**(1) (1992), 85–93.
- [214] P Zinn-Justin, *The six-vertex model on random lattices*, Europhys. Lett. **50**(1) (2000), 15–21.



Minerva Access is the Institutional Repository of The University of Melbourne

**Author/s:**

Elvey Price, Andrew

**Title:**

Selected problems in enumerative combinatorics: permutation classes, random walks and planar maps

**Date:**

2018

**Persistent Link:**

<http://hdl.handle.net/11343/219277>

**Terms and Conditions:**

Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.